From process understanding to process control: Application of PAT on the cultivation of *Bordetella pertussis* for a whole cell vaccine

Mathieu Streefland

Promotor
Prof.dr.ir. J. Tramper
Hoogleraar in de bioprocestechnologie, Wageningen Universiteit, Nederland

Co-promotoren
Dr.ir. D.E. Martens
Universitair docent sectie bioproceskunde, Wageningen Universiteit, Nederland
Dr. E.C. Beuvery
Zelfstandig ondernemer, PAT Consultancy, Vianen, Nederland
Dr. L.A. van der Pol
Hoofd afdeling procesontwikkeling, Nederlands Vaccin Instituut, Bilthoven, Nederland

Promotiecommissie
Prof.dr.ir. J. van der Oost
Wageningen Universiteit, Nederland
Prof.dr.ir. G. van Straten
Wageningen Universiteit, Nederland
Dr.ir. G. Zijlstra
DSM Biologics, Groningen, Nederland
Prof.dr. C.J.P. Boog
Nederlands Vaccin Instituut, Bilthoven, Nederland; Universiteit van Utrecht, Nederland

Dit onderzoek werd uitgevoerd binnen de onderzoekschool VLAG.

From process understanding to process control: Application of PAT on the cultivation of *Bordetella pertussis* for a whole cell vaccine

Mathieu Streefland

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit
Prof.dr. M.J. Kropff
in het openbaar te verdedigen
op vrijdag 20 maart 2009
des namiddags te half twee in de Aula

From Process Understanding to Process Control: Application of PAT on the cultivation of *Bordetella pertussis* for the development of a whole cell vaccine

Mathieu Streefland

PhD thesis, Wageningen University, The Netherlands, 2009
With summaries in English and Dutch

ISBN: 978-90-8585-357-2

Voor Mariska
voor Bouke

Contents

Chapter 1

Introduction

1

In the wake of the PAT (Process Analytical Technology) initiative launched by the American FDA early 2003, the Netherlands Vaccine Institute has started the Parametric Release, or PaRel project. Together with project partners Siemens Pharma (Zwijndrecht, Belgium) and Applikon Biotechnology (Schiedam, The Netherlands) a new approach towards the development of modern manufacturing processes for vaccines was investigated. The project ran from 2003 until the end of 2007 and its outcomes were used for the writing of this thesis. New analytical methods for both the product and the process are introduced for the cultivation process step for a whole cell vaccine against whooping cough disease, one of the oldest approved biopharmaceutical products. The information from these new methods, combined with the long manufacturing history and process knowledge already available at the Netherlands Vaccine Institute, have resulted in a science-based rationale for the cultivation of *Bordetella pertussis* bacteria for the preparation of this vaccine.

Bordetella pertussis and the whole cell vaccine against whooping cough disease

Bordetella pertussis is the causative agent of whooping cough disease. It is a gram negative, obligate aerobic cocco-bacillus, which means it is shaped as a short rod. It is closely related to the other species in the *Bordetella* genus, *B. parapertussis* and *B. bronchoseptica*, which are also pathogens. *B. pertussis* is the only exclusively human pathogen, while the others have a broader host range.

The whole cell vaccine against whooping cough disease consists of inactivated bacteria. This means that the production process is straightforward. After cultivation, the cells are concentrated using continuous centrifugation and then inactivated at 56 °C. The bulk vaccine is then usually adjuvated using aluminium phosphate or aluminium hydroxide and mixed with the vaccines against diphteria, tetanus and polio to constitute the DTP-polio vaccines that are commonly used for infant vaccination world wide.

The whole cell pertussis component is currently replaced by an acellular variant in most vaccine preparations, because of the adverse effects caused by cellular components of the *B. pertussis* bacterium. These acellular vaccines usually consist of a mixture of several virulence factors of *B. pertussis*: pertussis toxin, pertactin, FHA, or fimbriae (see Fig. 1). Several formulations are on the market with one or more of these proteins. These virulence factors are all present on the outer membrane of *B. pertussis*, or, like pertussis toxin, excreted into the culture medium. These proteins can induce a protective immune response after vaccination, some by themselves (pertussis toxin), others in combination with each other. The exact mechanism for induction of protective immunity through vaccination is still largely unclear. This means that vaccination with a whole cell vaccine may intrinsically be better than vaccination with an acellular vaccine, because it presents the same components to the immune system as a natural infection and is therefore able to induce a more “complete”

immune response[1]. For now, the safety concerns in the whole cell vaccine debate have won over the efficacy benefits, which means that from 2005 the Netherlands have also adopted an acellular vaccine against whooping cough.

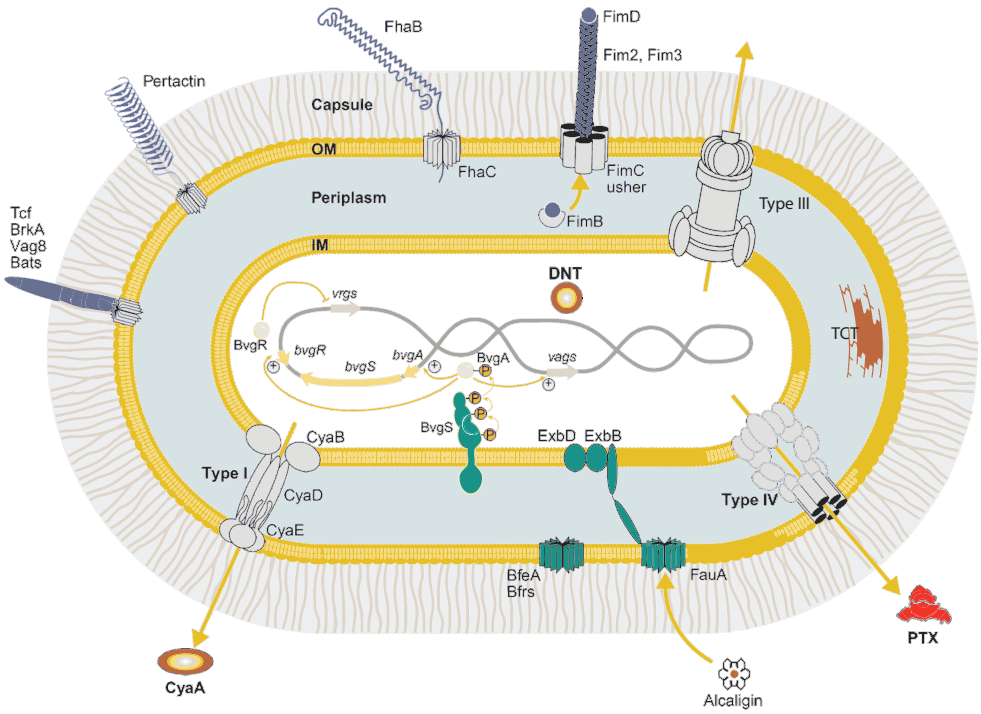


Figure 1. Schematic representation of the *Bordetella pertussis* bacterium. The main virulence factors are shown as well as the bvg regulon system that controls the expression of most of these virulence factors. This drawing is taken from Locht et al. (2001) Curr Opin. Microbiol. 4:89-892.

PAT for biopharmaceuticals

Being second only to the aviation industry in terms of regulatory burden, the pharmaceutical industry is dominated by strict compliance to safety and quality rules, such as those prescribed in the Good Manufacturing Practice (GMP) regulatory body. These regulations have resulted in an industry, which heavily relies on “quality by testing” instead of “quality by design”. In other words, product quality and safety in manufacturing are assured through protocols (Standard Operating Procedures or SOP’s) and end-product testing. Process changes or optimizations require extensive (clinical) testing and additional paperwork in order to comply to regulation. Therefore process improvement is often not pursued and many of the established processes are outdated or sub-optimal.

In other industries, such as the petrochemical industry, on line monitoring of product quality and continuous process optimization strategies are already common. The American Food and Drug Administration (FDA) recently recognized the need for the pharmaceutical industry to come up to the same level. For this, the FDA launched its Process Analytical Technology (PAT) initiative[3]. In summary, it prescribes quality assurance to become part of the process (i.e. quality by design) rather than relying on final product testing at the end of manufacturing (quality by testing). This means that processes need to be well characterized and the critical attributes of the process, the product and their interaction need to be known. In short, when one can not use final product testing as a means of quality assurance, the processes themselves need to be able to yield product of consistent high quality, without the necessity of final product testing as a fail safe.

The European Medicines Agency (EMEA) has adopted the PAT principles as well as the Japanese authorities. This resulted at the International Conference for Harmonization in the writing of three new guidance documents (Q8, Q9 and Q10), which can be considered the world standard for this new regulatory concept for product and process quality[4]. For pharmaceutical small molecules with relatively simple production processes, the application of PAT is becoming more and more common. For biological products acceptation of the PAT concepts has been much slower. This led the FDA to invite biopharmaceutical companies to work with them for pilot submissions of biopharmaceutical PAT applications[5].

The PaRel project was aimed at implementation of PAT on the *cultivation* process step of a whole cell vaccine against whooping cough disease. This process step involves the batch cultivation of the *Bordetella pertussis* bacterium. Because the cultivation process step is the most complicated and critical step in this production process, it was chosen for the development of the tools, equipment and knowledge necessary for full PAT application. Compared to a small molecule drug, this vaccine can be considered as relatively undefined and complex. This brings new challenges in how to assure product quality during processing. The FDA has acknowledged to project partner Siemens, that the approach chosen in the PaRel project can lead to approval of a PAT application for a complex biopharmaceutical product like a whole cell vaccine.

Contents of this thesis

Understanding of the cultivation process involves understanding of the interior physiology of the cell and how it responds to changes outside the cell, either due to process changes or disturbances. Especially in a batch-wise operated bioreactor, this understanding is important, because a batch cultivation process typically has continuously changing conditions, wich can influence cell physiology. In order to achieve this, a full genome DNA microarray analysis was developed for monitoring of mRNA expression profiles or *Bordetella pertussis*. Changes in the environment of a cell usually result in changes inside the cell. A major part of these intracellular effects can be identified with a microarray analysis. By measuring the gene expression levels during cultivation, the pathways or specific proteins that are affected by changes during processing can be identified. In this way, disturbances of specific process parameters can be assessed for impact on product quality. Also the reproducibility between batches can be investigated and the optimal harvest point can be determined by examining the gene expression profile.

Some genes are more important than others for the efficacy of the vaccine, e.g. to induce a protective immune response after administration. In a whole cell vaccine, the outer membrane proteins are the targets that are presented to the immune system, therefore the outer membrane protein composition is considered to be of critical importance for vaccine quality (Fig 1). Identification of the genes that are important for the correct outer membrane composition in terms of inducing a protective immune response is the subject of Chapter 2. These are mostly genes that are involved in the virulence of *Bordetella pertussis*, meaning that they play a role in the colonization and infection of the human host.
The virulence genes of the *Bordetella* genus are regulated by a highly conserved molecular switch, called the *bvg* system (Fig. 1). This means that a single extracellular signal can induce the complete physiological switch between the virulent and the non-virulent state[6]. All genes involved are under the control of the same operon system. For some strains these virulence genes were already investigated[7, 8], but not in terms of optimal outer membrane composition for vaccine manufacturing.

We investigated the genes in the NVI's vaccine strain *B. pertussis 509* and compared the genes that are differentially regulated between the virulent and non-virulent state with published results from other strains[7]. Fifty-six marker genes were identified to be conservatively regulated between strains. These genes are important for virulence and can be expected to be involved in inducing a protective immune response[9]. Microarray analysis of the expression levels of these marker genes of samples taken during processing could then be used to investigate the impact of process disturbances or failures on the expected outer membrane composition.

One of the process parameters that was investigated is the dissolved oxygen concentration (Chapter 3). Because *B. pertussis* is an obligate aerobic organism, the availability of oxygen during cultivation is essential. To investigate this, a series of cultivations were oxygen limited for 90 minutes. Samples were taken immediately before and after limitation and at the end of cultivation for microarray analysis. Oxygen limitation had a strong effect on gene expression levels of many genes immediately after the event, including virulence genes. At the end of cultivation, however, the oxygen limited cultures could not be distinguished from the standard control cultivations. This indicates that oxygen limitation can have an effect on vaccine quality, but this effect is reversible. As long as oxygen limitation does not occur at or near the end of cultivation it is not a critical factor for the manufacture of the vaccine.

Using microarrays for the monitoring of the crucial genes allowed for the analysis of the critical process attributes for the cultivation of *B. pertussis*. High expression of the 56 virulence marker genes is associated with high cellular virulence and thus good expected product quality, while low expression levels are associated with poor expected quality. Therefore, a weighed average of the expression levels of these virulence marker genes can serve as a predictor for the quality of the bacterial suspension at the end of cultivation. This so called product quality score can be used to compare batches or samples of the same batch with each other in an objective manner.

The product quality score was used to determine one of the most critical events in any batch cultivation: the optimal harvest point (Chapter 4). As biomass accumulates exponentially, nutrient concentrations drop exponentially (Fig. 2). Limitation of nutrients can result in a range of unwanted effects including cell lysis or the suppression of virulence genes[10]. In order to investigate the optimal harvest point, four identically operated cultivations were each sampled at 11 time points for microarray analysis. A continuously changing gene expression pattern over time was expected, because of the continuously changing extracellular environment in a batch cultivation. However, gene expression proved to be relatively constant, which resulted in high product quality scores during most of the cultivation. Towards the end of cultivation, the product quality score dropped sharply (Fig. 2). This coincided with the depletion of the nutrients lactate and glutamate. This pattern allowed accurate determination of the optimal harvest point.
By measuring lactate and glutamate concentrations during cultivation, the bacteria can now consistently be harvested before nutrients become limiting, assuring the optimal composition of the bacterial outer membrane.

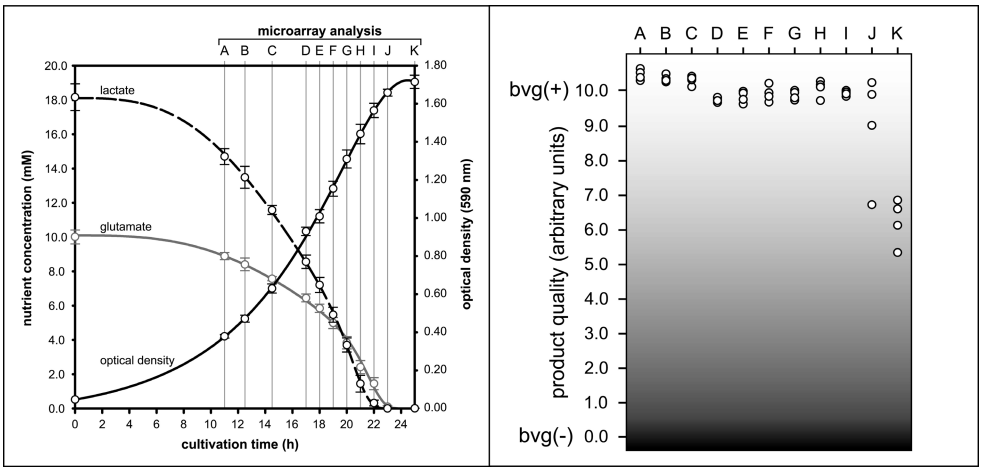


Figure 2. Batch profile analysis for the cultivation of *B. pertussis* (taken from Chapter 4)**.** The left diagram shows the average growth curve of four cultivations going up and the average nutrient concentrations (lactate and glutamate) going down. A to K mark the points at wich samples for microarray analysis were taken. The right diagram shows the relative product quality scores at the sample points A to K. It is clearly shown that the product quality score is consistently high for most of the process. Towards the end of the cultivation (points J and K) a sharp drop in the score is shown. This corresponds with the depletion of important nutrients (left diagram).

The definition of PAT states that it is a "*system for designing, analyzing and controlling manufacturing through timely measurements of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality*." With the at line measurement of nutrient concentrations and the supporting evidence that this correlates with virulence gene expression and subsequently outer membrane composition, the requirements for a PAT application are fulfilled, at least for the harvest point determination. For other critical process attributes several sensors are available (i.e. for pH or dissolved oxygen). However, these sensors only monitor a single parameter, which generates the risk of either missing a critical parameter or having to mount numerous probes on the bioreactor system for each individual parameter. To be able to monitor many parameters simultaneously, a near infrared (nIR) spectroscopic probe was introduced. This probe was placed inside the bioreactor and measured a spectrum between 800 and 2500 nm. This region, situated between visible light and far infrared, is sensitive both for chemical changes (band vibration overtones) and physical changes (combination bands), such as optical density, viscosity, particle size and particle morphology. This makes this technique highly suitable for monitoring bacterial cultivation processes as was confirmed by initial pilot studies (Chapter 5).

To control the manufacturing process through these "timely measurements" (as stated in the PAT definition) the on line process data, including nIR, needs to be readily available for process control models that feed back into the process. To allow this, project partner Siemens developed the SIPAT software. This software gathers all data that is measured online (i.e. pH, dissolved oxygen, temperature, nIR, controller outputs, gas flows, etc) and stores them in a central database with aligned timestamps. This database is accessible for process models that can be run in the integrated Umetrics Simca software. Output from these models flows back into the bioreactor control system in order to make adjustments to the process. This software allows the use of any process sensor for the monitoring and control of any process step. In this case, the development of the SIPAT software allowed the integration of nIR data with the other process data, making a true PAT application possible.

With the PAT initiative came also the concept of process design space. The International Conference for Harmonization (ICH) defines the design space as "the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality"[11]. This means that product quality is no longer assured at specific process settings, but rather a range of settings is explored. One of the best ways to explore the process design space is by using Design of Experiments (DoE)[12]. In this way a number of critical process parameters is tested at a range of settings using a minimal number of experiments. The main benefit of this approach is that the interaction between the critical parameters at different settings is investigated. We used this approach to explore the process design space of the cultivation of *B. pertussis* by executing a series of designed experiments in which several critical process parameters were varied simultaneously (Chapter 6).

Based on these experiments, a process model was constructed that describes the design space for the cultivation of *B. pertussis* for the tested ranges of the critical parameters. This process model can be executed on line using the SIPAT software, which allows an on line check if a new process is running within the tested process design space. Newly gathered data can be added to the process model so that it becomes more accurate over time. Ultimately this model can be validated so that it assures product quality on line, allowing real time product release.

The work in this thesis demonstrates that although the initial PAT guidance was only intended for small molecule drugs, the same principles can also be applied to biopharmaceutical products. Biopharmaceutical processes, and especially the cultivation process step, are intrinsically more complex than processes for chemical drugs. By applying a sound scientific approach for development of process understanding and by using the appropriate sensors and monitoring techniques, a biopharmaceutical process can become as stable and capable

as any “world class” manufacturing process (see Chapter 7).

The ultimate benefit of PAT application will be mitigation of risks (failures) during manufacturing, flexibility in process optimization and safer and better products. The effort taken for building process understanding will be rewarded when scale up, process changes or even medium optimizations are more easily implemented throughout the product’s manufacturing lifetime. Furthermore, on line monitoring of product quality and subsequent adjustment to the optimal trajectory will lead to less failed runs and a more consistent product of high quality. Real time quality assurance will lead to shorter cycle times and lesser stockpiles. These benefits have made the FDA decide to make PAT application the mandatory backbone of future pharmaceutical process development and manufacturing. This thesis demonstrates that it can even be applied on a complex and undefined product such as a whole cell vaccine, which means it should also be feasible for any (bio)pharmaceutical product currently on the market. The PAT initiative can therefore help to increase the safety and efficacy of medicines in general while reducing the time to market for new products and the operational costs of manufacturing. This is good news, for the regulatory inpectors, the manufacturers and the patients.

References

1. Locht, C., (2001) Cellular and acellular anti-pertussis vaccines. *Annales pharmaceutiques francaises* 59, 198-205
2. Locht, C., *et al*. (2001) Bordetella pertussis, molecular pathogenesis under multiple aspects. *Current opinion in microbiology* 4, 82-89
3. FDA (2004) Guidance for Industry, PAT - a framework for innovative pharmaceutical development, manufacturing and quality assurance. http://www.fda.gov/cder/guidance/6419fnl.pdf
4. ICH (2008) Quality Guidances. http://www.ich.org/cache/compo/363-272-1.html
5. FDA (2008) Submission of Quality Information for Biotechnology Products in the Office of Biotechnology Products; Notice of Pilot Program. *Federal Register* 73, 37972-37974
6. Smith, A.M., *et al*. (2001) The virulence factors of Bordetella pertussis: a matter of control. *FEMS Microbiol. Rev.* 25, 309-333
7. Cummings, C.A., *et al*. (2006) Species- and strain-specific control of a complex, flexible regulon by Bordetella BvgAS. *J.Bacteriol.* 188, 1775-1785
8. Hot, D., *et al*. (2003) Differential modulation of Bordetella pertussis virulence genes as evidenced by DNA microarray analysis. *Mol.Genet.Genomics* 269, 475-486
9. Hamstra, H.J., *et al*. (1995) The purification and protective capacity of Bordetella pertussis outer membrane proteins. *Vaccine* 13, 747-752
10. Nakamura, M.M., *et al*. (2006) Growth phase- and nutrient limitation-associated transcript abundance regulation in Bordetella pertussis. *Infect.Immun.* 74, 5537-5548
11. ICH (2005) Pharmaceutical Development (Q8). *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use* http://www.ich.org/LOB/media/MEDIA1707.pdf
12. Montgomery, D.C. (2001) Experimental Design for Product and Process Design and Development. *Journal of the Royal Statistical Society* 48, 159-177

CHAPTER 2

PAT for vaccines: The first stage of PAT implementation for development of a well-defined whole-cell vaccine against whooping cough disease

Mathieu Streefland[1], Bas van de Waterbeemd[1], Hester Happé[2], Leo A. van der Pol[1], E. Coen Beuvery[3], Johannes Tramper[4], Dirk E. Martens[4]

[1]Netherlands Vaccine Institute (NVI), Unit Research and Development, PO Box 457, 3720 AL Bilthoven, The Netherlands. [2]Leiden University Medical Center (LUMC), Department of Human Genetics, PO Box 9600, 2300 RC Leiden, The Netherlands. [3]PAT Consultancy, Kerkstraat 66, 4132 BG Vianen, The Netherlands. [4]Wageningen University, Food and Bioprocess Engineering Group, PO Box 8129, 6700 EV Wageningen, The Netherlands

Vaccine (2007) 25:2994-3000

Supplementary data can be found with the on line published version of this chapter at the journal's home page

2

Abstract

Since variation in process time and process output is commonly accepted to be inevitable for biological processes, application of Process Analytical Technologies (PAT) on these processes is challenging. In this paper the applicability of PAT on the cultivation of *Bordetella pertussis* bacteria as part of the manufacture of a vaccine against whooping cough disease is investigated. Scrutinizing and eliminating the most prominent sources of variance make the cultivation process step highly reproducible. Furthermore, the use of DNA microarrays allows investigation of how disturbances influence cellular physiology and product quality. Marker genes for product quality were identified, providing the means to quantitatively assess product quality, which is hardly possible using the mandatory animal tests for product quality. The tools and results described in this paper, combined with suitable on line measurements, can make full PAT application for this process step possible. Ultimately, the process can be designed and controlled towards consistent end product quality.

Introduction

Production processes of biologicals are generally considered less controllable and more susceptible to variation than processes for chemical drugs due to poor process understanding. Vaccines are considered to be at the low end of this spectrum. The production of vaccines has historically only been considered viable as a low-cost, low-tech industry and this became even more difficult after the introduction of GMP in 1980. Despite all the breakthroughs in immunological sciences of the past decades, the production processes for vaccines are essentially unchanged. Only recently vaccines have attracted the interest of large pharmaceutical companies for being potential block-buster candidates[1]. This means that vaccines are more and more becoming high-value biologicals. This generates a drive for more robust and efficient manufacturing processes for vaccines.

The US Food and Drug Administration (FDA) has recognized the need for robust and efficient manufacturing throughout the pharmaceutical industry. The high regulatory burden in this industry prevents continuous process optimization and process adaptations as a means for increasing manufacturing efficiency. In order to overcome these regulatory barriers, the FDA has launched the Process Analytical Technology (PAT) Initiative in 2003[2].

PAT is considered by the FDA to be the *desired future state of pharmaceutical manufacturing*. It requires a risk-based approach in which the critical process parameters or attributes are assessed for their risk for product quality. Monitoring and control of these parameters can assure good process performance and even the quality of the product as it is formed[3]. This requires an on line monitoring and control system that can measure the critical parameters during processing and take appropriate action upon any deviation. Although the application of PAT in a vaccine manufacturing process may seem challenging, a science-based approach has much to win in this industry, since production processes for vaccines are usually highly empirical. Before a PAT system can be designed, the process needs to be sufficiently understood and the critical process attributes that need to be monitored and preferably be controlled on line have to be identified. In this paper these first stages of PAT application are investigated for the cultivation process step of a vaccine against whooping cough disease.

For inactivated or attenuated live vaccines product quality is primarily determined during cultivation of the actual pathogen. For a-cellular or subunit vaccines this might be the cultivation of the host cells (i.e. CHO cells or *E. coli* bacteria), which produce the required vaccine components. The focus of this work is on a whole-cell vaccine against whooping cough disease. This vaccine is based on a heat-inactivated suspension of the pathogen, the *Bordetella pertussis* bacterium. Especially with such an inactivated whole-cell vaccine, the cultivation process step is the most relevant process step for PAT application, because in this step product quality is mainly determined.

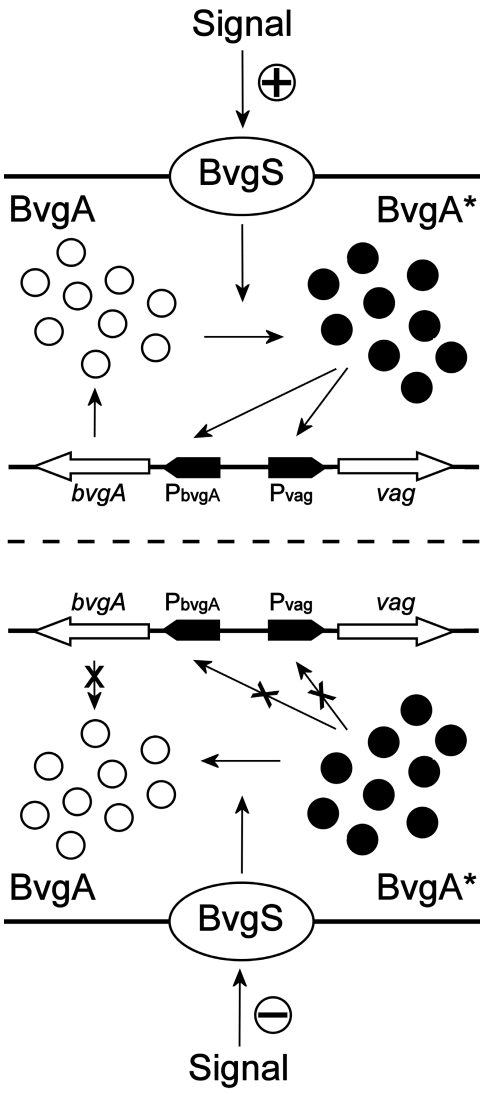


Figure 1. Schematical representation of the regulatory BvgA/S network of B. pertussis. Virulence inducing conditions (top half) trigger BvgS to phosphorylate BvgA through a phosphorelay pathway. Activated BvgA (BvgA*) acts as a transcription factor and binds to the promotor sequences (PbvgA, Pvag) of Bvg-activated genes to induce transcription. Because BvgA and BvgS are Bvg-activated genes as well, a positive feedback loop exists that results in high abundance of activated BvgA. Activated BvgA also binds to promotor sequences of Bvg-repressed genes to inhibit transcription (not shown). Virulence repressing conditions (bottom half) result in a de-phosphorylation of BvgA to stop the induction of Bvg activated genes. Expression of Bvg-repressed genes will no longer be inhibited and these genes will be subsequently activated (not shown). (Adapted from Figure 5 of Scarlato *et al*[10]).

A whole-cell vaccine might in principal provide a broader immune response then an a-cellular vaccine, because it inherently contains all necessary components for inducing protection. Several studies have indicated that whole-cell vaccines may have a better efficacy than a-cellular vaccines[4-6], but the side effects after vaccination have resulted in a strong decrease of their application. These side effects (i.e. swelling, fever) are mainly caused by lipopolysaccharide (LPS). The presence of LPS anchored in the outer membrane is relatively constant under normal production conditions[7], whereas the expression of the outer-

membrane proteins that are important for inducing a protective immune response can vary dependending on cultivation conditions[8]. Optimal outer-membrane protein composition can therefore be considered a critical attribute for product quality. More precisely, the expression levels of virulence factors are critical for product performance.

The mechanism through which *B. pertussis* regulates the expression of virulence factors has been extensively studied[8-10]. Most known virulence factors are under control of a two-component regulatory system named BvgA/S. This system functions as a molecular switch that acts to activate or repress virulence factor expression depending on environmental conditions (Figure 1). In order to obtain a high quality product, this switch needs to be in the bvg mode, resulting in bacteria that express virulence factors abundantly.

Having defined the properties of the product that are critical for product quality, accurate measurement of these properties is the next step for PAT application. This requires much more precise information than is provided by the mandatory animal tests that are currently used for this vaccine[11]. DNA microarrays for analysis of gene expression levels and profiles are proposed for assessment of product quality at the end of cultivation. mRNA expression levels are expected to serve as an indicator for the expression of virulence factors and other proteins relevant for prediction of product quality and process performance.

The availability of the full genome sequence of *B. pertussis*[12], created the opportunity to design a DNA microarray covering 93% of the *B. pertussis* open reading frames. Recent research has shown that the use of DNA microarrays can accurately determine the virulence state of the bacterium[13;14]. This technique can be applied in several manners. First, the genes crucial for virulence, and thus product quality, can be assessed. Second, overall expression levels at the end of cultivation can be compared with other batch runs. Good correlation between overall mRNA expression levels will indicate comparable end-product quality, whereas disturbances in expression levels can be identified and investigated.
After the desired product is defined and its quality can be accurately assessed, the biggest challenge that remains is the reproducible performance of the actual manufacturing process. The ability to do this cultivation in a robust and reproducible manner is a prerequisite for implementation of PAT, as it demonstrates process understanding and the ability to control critical process variance. Bacterial cultivation has many sources of variance, some of which could risk end quality to be out of specifications. Assessment of all relevant sources of variance and subsequent assessment of their potential risk for end-product quality, although complicated, could be one of the most rewarding exercises in the PAT trajectory, because improvement of process reproducibility is of key importance for robust processing.

In this paper we demonstrate the first stages of PAT application in the cultivation process of *B. pertussis* bacteria for a vaccine against whooping cough disease. First, the elimination of the most obvious causes of variation reduces the variation of the cultivation process. Next the use of DNA microarrays for identification of the relevant genes affecting product quality is demonstrated. Finally, we show how DNA arrays can be used to check the reproducibility of the cultivation with respect to the entire transcriptome and the relevant marker genes.

Materials and methods

Bacterial strain, media and growth conditions

The *B. pertussis* 509 strain (Netherlands Vaccine Institute, Bilthoven, The Netherlands) used in this study was collected in 1963 as a clinical isolate and used as vaccine strain in the combined diphtheria, pertussis, tetanus and polio (DPTPolio) vaccine until 2005.

All bacterial cultures were grown in chemically defined THIJS-medium[15], which consists of basic medium and supplement. Supplement (1% v/v) was added to the basic medium shortly before inoculation. The pH after addition of supplement was 7.1. The preculture was inoculated with 10ml *B. pertussis* 509 working seedlot (cells at OD_{590}=1.0±0.05 frozen with 10% glycerol, stored at -140 °C) and grown in 500ml shake flasks containing 200ml medium at 35°C, on an orbital shaker at 200 RPM. When the starting culture reached OD_{590}=1.0±0.05, it was used to inoculate either a bioreactor (reproducibility experiments, cultivation A through E), or other shake flasks (Bvg-modulation experiments). Start density was OD_{590} = 0.050±0.005 for all cultivations.

Bioreactor cultivations

All bioreactor experiments were carried out in a 7 L in situ sterilizable glass bioreactor with a working volume of 4 L. Temperature, dissolved oxygen (DO) concentration, pH and stirrer speed were controlled at 37°C, 30%, 7.2 and 450-650 rpm, respectively. Oxygen concentration was first controlled with increments of 50 rpm in stirrer speed until 650 rpm was reached. After this setpoint was reached, DO was controlled by increasing the fraction of oxygen in the headspace. After the bioreactors were filled with 4L THIJS medium and reached setpoint temperature and DO=100%, they were inoculated with 200 ml preculture at OD_{590}=1.0±0.05. Samples were taken at time points indicated in for OD_{590} and at line nutrient-concentration measurements. Samples were sterile filtered (0.22 µm) and stored at -20 °C for later NMR analyses. Growth data was fitted against a standard asymmetrical sigmoid curve ($y=a+b/(1+\exp(-(x-d\ln(2^{1/e}-1)-c)/d))^e$) using Tablecurve 2D software (AISN Software, USA).

Nutrient-concentration analysis

Nutrient (lactate and glutamate) concentrations were determined at line using an YSI 2700 analyzer (Yellow Springs Instruments, Yellow Springs, USA). Supernatants were later analyzed by ^{1}H-NMR using a JEOL JNM ECP 400 spectrometer operating at 400 MHz (JEOL, Tokyo, Japan) equipped with a JEOL Stacman autosampler for 16 samples. Supernatants were analyzed by adding 0.1 ml of D2O containing 3-(trimethylsilyl)[D4]proprionic acid sodium salt (TMSP, 0.167 mM) to a 0.9 ml sample. The water signal was suppressed by irradiating the signal with standard NMR software. The spectra were referenced using the TMSP signal at 0 ppm. Lactate and glutamate concentrations were quantified by integration of the relevant signals. NMR was also used to check samples for any waste metabolites.

Virulence modulation

Virulence was repressed with addition of either 50mM $MgSO_4$ or 20mM niacin to the basic medium. The acidity of niacin was compensated using additional NaOH (5M). To maintain Na^+ concentration at 75mM in all media, the amount of NaCl was adjusted accordingly. When the preculture reached OD_{590}=1.0±0.05, 10ml suspension was used to inoculate each of the 12 experimental shake flasks (4 normal medium, 4 $MgSO_4$ containing medium and 4 niacin containing medium). Harvest samples were analyzed with DNA microarrays.

Microarray design

Samples were analyzed using a full genome *B. pertussis* microarray. Based on the complete genome sequence of *B. pertussis* Tohama I[12], a set of 3582 70-mer oligonucleotides was developed at Operon (Cologne, Germany), covering 93% of all open reading frames. Oligonucleotide pellets were dissolved in 50% DMSO (v/v in water) to a concentration of 20μM and spotted in triplicate on UltraGAPS II coated slides (Corning, New York U.S.A.), together with 420 control spots, using the Omnigrid 100 microarray spotter (GeneMachines, San Carlos U.S.A.).

RNA isolation

For fixation of the RNA expression profile, 1 volume bacterial culture was mixed with 2 volumes RNAse retarding solution[16]. For each microarray sample 2.5 ml cultivation at an OD_{590}=1.0 was used. Samples at other optical densities, were adjusted accordingly, so that an equal amount of cells was used for RNA isolation. The samples were concentrated by centrifugation and treated with Tris-EDTA buffer, containing 0.5mg/ml lysozyme (Sigma-Aldrich, Zwijndrecht, The Netherlands) for 3 minutes. Total RNA was extracted with the SV Total RNA isolation system (Promega Benelux, Leiden, The Netherlands) according to manufacturer's protocol. Nucleic-acid concentration was adjusted by precipitation and spectral analysis was used to determine final nucleic-acid concentration and purity. RNA integrity was confirmed with the Bioanalyzer RNA6000 Nano assay (Agilent Technologies, Amstelveen, The Netherlands), according to the manufacturer's protocol.

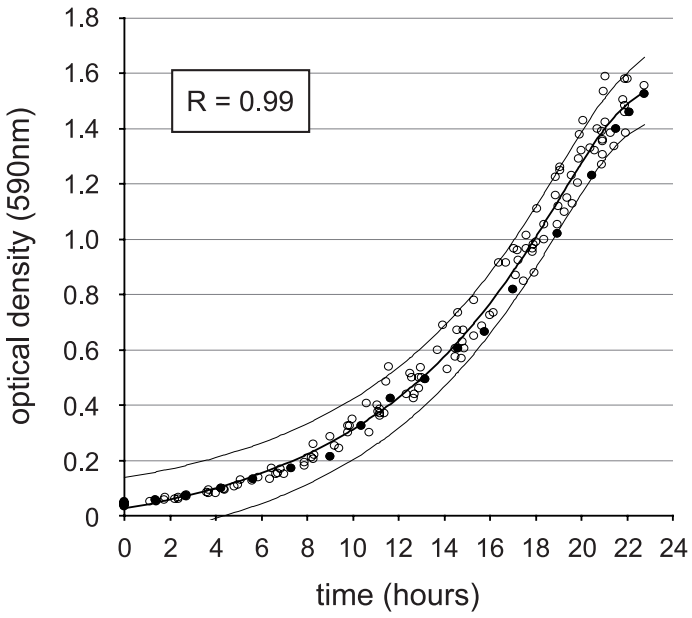


Figure 2. Open circles (○) indicate growth curves (OD_{590}) of 11 bioreactor cultivations. The center line indicates the fitted standard sigmoid function (R^2=0.99) and the outer lines indicate 95% prediction interval. Closed circles (●) show the growth data of cultivation C (Figure 4).

cDNA labeling and hybridization reactions

Total RNA from all experimental samples was reverse transcribed to cDNA and labeled with Cy3/Cy5 dyes using the Cyscribe Post-Labeling kit (Amersham Biosciences Benelux, Roosendaal, The Netherlands) according to manufacturer's protocol, with one deviation: only 2μl random nonamer primer and no oligo-dT primer was used per reaction to reverse transcribe 20μg total RNA. The labeled and purified cDNA samples were pooled in Cy3/Cy5 pairs and volumes were adjusted to 25μl by evaporation under low pressure. 25μl hybridization buffer was added, to a final concentration of 25% formamide, 5x SSC and 0.1% SDS. Samples were applied to the microarray slides and placed in a hybridization chamber (GeneMachines, San Carlos, U.S.A.) for 16-20 hours at 42°C in the dark.

Data acquisition and analysis

The microarrays were scanned with a ScanArray Express microarray scanner (Perkin Elmer, Groningen, The Netherlands) and median fluorescence intensities were quantified for each spot using ArrayVision software (Imaging Research, Roosendaal, The Netherlands). The

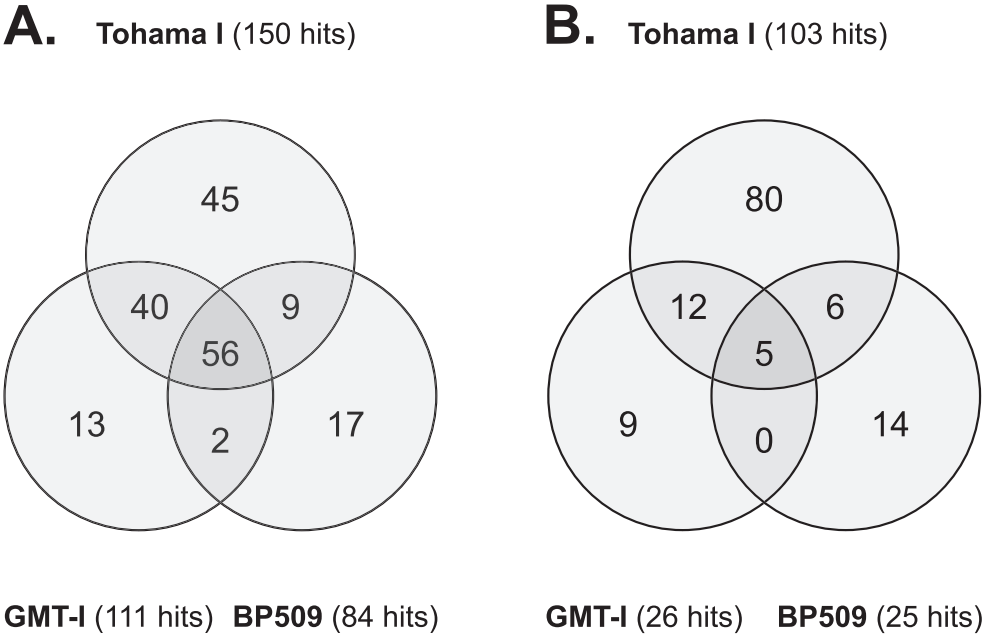


Figure 3. Comparison of virulence activated and repressed genes between the Tohama, GMT-1 and 509 strains of *B. pertussis*. Figure 3A shows a Venn diagram of the Bvg activated genes of three strains of *B. pertussis*. Numbers indicate the number of genes that are activated in virulent conditions and repressed in a-virulent conditions. Figure 3B shows a Venn diagram of the Bvg repressed genes of three strains of *B. pertussis*. Numbers indicate the number of genes that are repressed in virulent conditions and activated in a-virulent conditions. The central areas in both panels show the number of genes that are part of the core regulon as proposed by Cummings et al.[13]. Data for Tohama and GMT-1 strains are adopted from this work.

expression data were natural-log transformed, quantile normalized, and values of replicate spots were averaged. These data processing steps were done with the free statistical software R (WU Wien, Austria), using an in-house developed script.

Gene-expression values of the Bvg modulation experiment were calculated by comparing 4 replicate normal samples to 4 replicate $MgSO_4$-modulated samples. P-values were adjusted for multiple testing by calculating the False Discovery Rate (FDR). Fold Ratio (FR)-values were expressed as natural log of the normalized $MgSO_4$ / normal signal ratio. The same procedure was applied for comparing the normal samples to 4 replicate niacin-modulated samples. Observed differential gene expression was considered biologically significant when mean FR-values were above 2.0 and FDR values were below 0.10 for both the $MgSO_4$ and the niacin samples when compared to the normal samples.

Differential gene-expression levels of the reproducibility experiment (bioreactor cultivation A through E) were obtained through comparison with a common reference sample, containing equal amounts of RNA from each cultivation. Any genes that showed a FR>2.0 compared to the expression levels in the other cultivations were considered deviating.

Results

Bacterial growth

The cultivation method using THIJS medium described above resulted in a highly reproducible cultivation. Comparison of 11 bioreactor runs shows a high correlation between runs (R^2=0.99) and low variance (SE=0.055) (Figure 2).

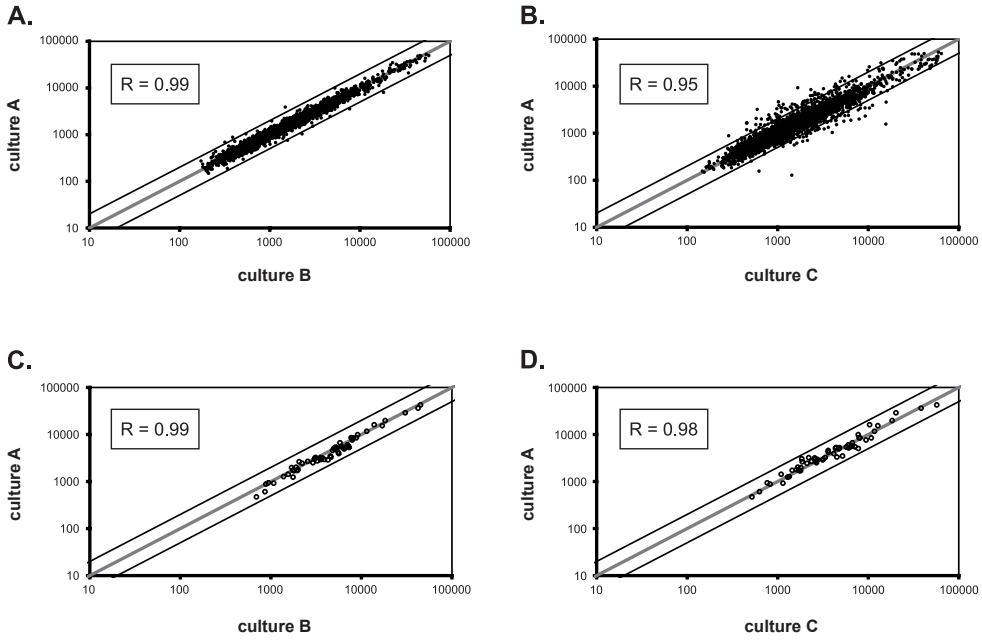


Figure 4. Comparison of gene expression levels at the end of bioreactor cultivation of *B. pertussis*. Center line indicates optimal correlation. Outer lines indicate 95% confidence interval. Figure 4a shows a full genome comparison of the gene expression levels of Culture A with Culture B. Closed circles (●) indicate relative gene expression levels. Correlation coefficient R=0.99. Figure 4B shows the full genome comparison of the gene expression levels of Culture A with Culture C. Closed circles (●) indicate relative gene expression levels. Correlation coefficient R=0.95. Figure 4C shows the comparison of the gene expression levels of the Bvg activated genes of Culture A with Culture B. Open circles (o) indicate relative gene expression levels. Correlation coefficient R=0.99. Figure 4D shows the comparison of the gene expression levels of the Bvg repressed genes of Culture A with Culture C. Open circles (o) indicate relative gene expression levels. Correlation coefficient R=0.98.

Genes involved in virulence

Chemical induction of an a-virulent state, using 50mM $MgSO_4$ or 20mM niacin, and subsequent comparison with normal cultivation conditions revealed the genes involved in virulence of the *B. pertussis* 509 strain. These genes were compared with the core regulon proposed for the *B. pertussis* GMT-1 and Tohama strains by Cummings[13]. Only genes that showed an effect in both modulating conditions ($MgSO_4$ and niacin) were considered a hit. Any hits that were specific for either condition (Figure 3) were not included. The center of the Venn diagrams shown in Figure 3 represents the number of genes of the *B. pertussis* 509 strain that overlap with the core regulon proposed by Cummings and co-workers[13]. A list of the genes and their expression levels corresponding with the data in Figure 3 can be found in an online supplementary data file at doi:10.1016/j.vaccine.2007.01.015.

Comparison of RNA expression profiles

Comparison of the overall gene expression levels between cultivations at the harvest point may be a relevant indication for the reproducibility. Five bioreactor cultivations (named A through E) were analyzed at harvest point with DNA microarrays. Four of these cultivations showed a good correlation between gene expression levels (R=0.98-0.99). Two typical cultivations are compared in Figure 4A. One cultivation (C) deviated from the others in overall expression profile and showed a lower correlation with the other cultivations (R=0.95-0.96), Figure 4B). Analysis of the process data revealed that cultivation C was harvested just after the main nutrients lactate and glutamate were (nearly) depleted from the medium (0.00 mM and 0.66 mM, respectively). The other four cultivations were harvested with 1 – 3 mM of both still present in the medium. Cultivation C did not show a deviation in bacterial growth (Figure 2) or other process data compared to the other batches. Investigation of the deviating genes (FR > 2.0) in cultivation C revealed that most are coding for ribosomal proteins, sulphate transport or ABC transporters, indicating decreased protein synthesis, changes in metabolism and general cellular activity. A table containing all deviating genes of cultivation C is included in the supplementary data file.

In contrast to the deviating overall gene expression levels, the genes of the core regulon correlate well for all cultivations (R=0.98-0.99), including cultivation C (Figure 4C and D). This indicates that the limitation of lactate at the point of harvest for cultivation C is not likely to result in a risk for product quality.

Discussion

Variation in process performance is a high risk for inconsistency of product quality. Identification of the factors that cause process variation will enable their control, which in turn will result in more robust and consistent processes. Biological processes such as the cultivation of bacteria are usually considered variable and hard to control. If, however, a bacterium is given only a limited number of degrees of freedom, it will adopt a predictable growth pattern, as demonstrated here. A chemically defined medium, consistent seedlot preparation and control of harvest points eliminate most of the variance for (pre)culture growth. The THIJS medium that is used for all cultivations[15,17] is designed to meet the organisms' nutritional demands at optimal growth conditions and allows little room for activation of other (waste) metabolic routes upon starvation or process disturbances, which is beneficial for cultivation reproducibility.

The use of DNA microarrays as a tool to monitor process consistency provides insight in the behavior of the organism during growth. Theoretically any disturbance that influences product quality should be reflected in the mRNA expression profile, since genes usually function in relation with other genes. Therefore, any disturbance or change should be reflected by alterations in the expression levels of at least several genes. We observed high correlations of both full genome and core regulon expression levels between cultivations (Figure 4).

This is indicative of high comparability at the point of harvest. Depletion of lactate and glutamate at the end of the cultivation resulted in a disturbance of the overall expression level in one cultivation (C), but the expression levels of the genes of the core regulon were not affected. Disturbances that affect the overall expression levels, but not the core regulon expression levels are not likely to be critical for product quality, since the virulent (bvg^+) state is not affected and the proteins important for vaccine efficacy are still expressed.

The genes involved in virulence of the *B. pertussis* 509 strain show considerable overlap with the genes that Cummings[13] found to be involved in virulence of the Tohama and GMT-I strains. Especially the most important and well-known Bvg-activated genes that code for protection inducing antigens (i.e. pertussis toxin, pertactin, filamentous haemagglutinin, see supplementary data file) seem to be conserved between all strains, while the Bvg-repressed genes are more variable. This supports the concept of the core and flexible regulon proposed by Cummings[13]. With these results we now have a set of marker genes for the virulence state of the bacteria. Because the Bvg-repressed genes show high variation between strains, it is hard to pinpoint the core regulon for virulence-repressed genes. However, for product efficacy the Bvg-activated genes are most relevant and evaluation of the virulence state of the product can be confined to the expression status of these genes only.

The genes that are involved in virulence of the *B. pertussis* 509 strain code for the most likely proteins that can induce an immune response. These results can therefore be a starting point for a reverse vaccinology approach to find new vaccine candidates[18].

DNA microarrays can be used for prediction of product quality in a discriminative way, while the usual mandatory efficacy tests only provide yes or no answers[11]. Before PAT can be applied on any process, products of poorer quality need to be distinguished from the higher quality products, even though both products would meet end-product quality demands. Only in this way can processes be optimized and made more robust in order to yield product of consistently high quality. Ultimately, mRNA expression data can be correlated to the signal of on line process monitoring sensors such as near infrared (nIR) or Raman spectroscopy to build a model that can assess process performance and predict product quality on line as is desired by the FDA in their PAT initiative.

Conclusions

Generally, the cultivation of vaccines is considered to have high inherent variance and is therefore not as suitable for PAT application as chemical (synthesis) or physical (blending) process steps. This paper demonstrates that elimination of the major sources of variance leaves a cultivation process step that is highly reproducible. Application of modern techniques such as DNA microarrays provides a tool for science-based evaluation of product quality. Disturbances in mRNA expression levels of specific genes provide good starting points for further research, because the physiological cellular response is visualized. The reproducible process and the tools to investigate disturbances and monitor consistency combined with a suitable on line monitoring tool (such as near infrared or Raman spectroscopy) would make PAT application on this process feasible. The discriminative data on product quality and process performance should be correlated to such on line measured data to build process models that control the process towards consistent end product quality. Product quality would become an inherent trade of the process.

Acknowledgements

The authors gratefully acknowledge Dr J. Pennings for his help with the microarray analysis and data processing, Dr B. Zomer for NMR measurements and Dr M.J. Vansteensel for help with the artwork. This work was partly funded by Senter Novem, The Netherlands, Grant Number TGSE3067.

Reference List

1. Sheridan C. The business of making vaccines. Nat Biotechnol 2005 Nov;23(11):1359-66.
   U.S.Department of Health and Human Services Office of Regulatory Affairs (ORA), Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Veterinary Medicine (CVM). Guidance for Industry1 PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. ww.fda.gov/cder/guidance/index.htm , 1-16. 1-9-2004. Ref Type: Internet Communication
3. Balboni ML. Process Analytical Technology Concepts and Principles. Pharmaceutical Technology 2003 Feb 10;(October 2003):54-66.
4. David S, van Furth R, Mooi FR. Efficacies of whole cell and acellular pertussis vaccines against Bordetella parapertussis in a mouse model. Vaccine 2004 May 7;22(15-16):1892-8.
5. Mills KH. Immunity to *Bordetella pertussis*. Microbes Infect 2001 Jul;3(8):655-77.
6. Olin P, Rasmussen F, Gustafsson L, Hallander HO, Heijbel H. Randomised controlled trial of two-component, three-component, and five-component acellular pertussis vaccines compared with whole-cell pertussis vaccine. Ad Hoc Group for the Study of Pertussis Vaccines. Lancet 1997 Nov 29;350(9091):1569-77.
7. van den Akker WM. Lipopolysaccharide expression within the genus Bordetella: influence of temperature and phase variation. Microbiology 1998 Jun;144 (Pt 6):1527-35.
8. Smith AM, Guzman CA, Walker MJ. The virulence factors of *Bordetella pertussis*: a matter of control. FEMS Microbiol Rev 2001 May;25(3):309-33.
9. Cotter PA, Jones AM. Phosphorelay control of virulence gene expression in Bordetella. Trends Microbiol 2003 Aug;11(8):367-73.
10. Scarlato V, Prugnola A, Arico B, Rappuoli R. Positive transcriptional feedback at the bvg locus controls expression of virulence factors in *Bordetella pertussis*. Proc Natl Acad Sci U S A 1990 Sep;87(17):6753-7.
11. IRWIN JO, STANDFAST AF. The intracerebral mouse-protection test for pertussis vaccines. J Hyg (Lond) 1957 Mar;55(1):50-80.
12. Parkhill J, Sebaihia M, Preston A, . Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat Genet 2003 Sep;35(1):32-40.
13. Cummings CA, Bootsma HJ, Relman DA, Miller JF. Species- and strain-specific control of a complex, flexible regulon by Bordetella BvgAS. J Bacteriol 2006 Mar;188(5):1775-85.
14. Hot D, Antoine R, Renauld-Mongenie G, . Differential modulation of *Bordetella pertussis* virulence genes as evidenced by DNA microarray analysis. Mol Genet Genomics 2003 Jul;269(4):475-86.
15. Thalen M, van den IJssel J, Jiskoot W, . Rational medium design for *Bordetella pertussis*: basic metabolism. J Biotechnol 1999 Oct 8;75(2-3):147-59.
16. Alberti S, Krause SM, Kretz O, . Neuronal migration in the murine rostral migratory stream requires serum response factor. Proc Natl Acad Sci U S A 2005 Apr 26;102(17):6148-53.
17. Thalen M, Venema M, van den IJssel J, . Effect of relevant culture parameters on Pertussis Toxin expression by *Bordetella pertussis*. Biologicals 2006 Sep;34(3):213-20.
18. Rappuoli R, Covacci A. Reverse vaccinology and genomics. Science 2003 Oct 24;302(5645):602.

CHAPTER 3

Evaluation of a critical process parameter: Oxygen limitation during cultivation has a fully reversible effect on gene expressions of *Bordetella pertussis*

Mathieu Streefland[1], Bas van de Waterbeemd[1], Joeri Kint[1,2], Leo A.van der Pol[1], E. Coen Beuvery[3], Johannes Tramper[2], Dirk E. Martens[2]

[1] Netherlands Vaccine Institute (NVI), Unit Research and Development, PO Box 457, 3720 AL Bilthoven, The Netherlands
[2] Wageningen University, Food and Bioprocess Engineering Group, PO Box 8129, 6700 EV Wageningen, The Netherlands
[3] PAT Consultancy, Kerkstraat 66, 4132 BG Vianen, The NetherlandsThe Netherlands.

Biotechnology and Bioengineering (2009) 102(1): 161-167
Supplementary data can be found with the on line published version of this chapter at the journal's home page

3

Abstract

Modern (bio)pharmaceutical process development requires thorough investigation of all process parameters that are critical to product quality. The impact of a disturbance of such a parameter during processing needs to be known so that a rational decision can be made about the release of the product. In cultivation processes the dissolved oxygen concentration (DO) is generally accepted as being a critical parameter. In this paper the impact of a 90 minute period of oxygen limitation during the cultivation of the strictly aerobic *Bordetella pertussis* bacterium is investigated. The cultivation is the most important process step for the manufacturing of a vaccine against whooping cough disease. Samples were taken immediately before and after oxygen limitation and at the end of cultivation of four oxygen limited and three control cultivations. DNA microarray analysis of the full transcriptome of the *Bordetella pertussis* bacterium revealed that a 90 minute period of oxygen limitation has a substantial effect on overall gene expression patterns. In total 104 genes were identified as a significant hit at any of the sample points, of which 58 were directly related to oxygen limitation. The other genes were mainly affected towards the end of cultivation. Of all genes involved in oxygen limitation none were identified to show a significant difference between the oxygen limited and control cultivations at the end of the batch. This indicates a fully reversible effect of oxygen limitation on gene expression. This finding has implications for the risk assessment of dissolved oxygen concentration as a critical process parameter.

Introduction

The manufacturing of vaccines, as well as all other pharmaceutical preparations, has to be executed within predefined specifications. When specifications have not been met, the batch or run cannot be released and has to be started again. However, sometimes little is known about the actual effect of the failure or deviation on product quality. When a certain disturbance causes the process to run outside the specifications (out of specs, OOS), the manufacturer has no choice but to reject the batch, regardless the actual effect of the disturbance on product quality. This means a manufacturer could very well reject a perfectly good batch because the process specifications have not been sufficiently supported by scientific evidence.

This undesirable situation has come to exist because the pharmaceutical industry is highly regulated and document controlled. In any other process industry the market demand to produce products of high quality on time and at low costs would have initiated research for better process understanding and process control. In the pharmaceutical industry, however, regulatory systems such as GMP have resulted in fixed processes that have to be controlled by extensive documentation such as SOP's, batch records and numerous other protocols. One might argue that a document with the correct signatures has become a replacement for actual product quality management.

In response to the continuously increasing paperwork, the American FDA launched the Process Analytical Technology (PAT) initiative in 2003 in order to make the pharmaceutical industry move away from document control and design processes that can be controlled towards a desired end-point quality. The PAT initiative is part of the much broader initiative "Pharmaceutical cGMPs for the 21st Century: a risk based approach" (see http://www.fda.gov) and has already been adopted by other authorities such as the European EMEA. Recently, new guidances released by the harmonization agency ICH (Q8, Q9, Q10) demonstrated a worldwide consensus regarding this subject. Generally, the FDA's definition is adopted which states: *PAT is a system for designing, analyzing and controlling manufacturing through timely measurements of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality.*

For each unit operation such a system needs to be developed and therefore, for each unit operation the parameters that are critical to process performance or product quality need to be known. For a document controlled industry like the pharmaceutical industry, this means a completely new way of product analysis and process development. This is especially true for the production processes of biologicals, or even vaccines, because these are generally considered harder to control and more variable than manufacturing processes for chemical drugs.

In this paper, we are investigating a PAT application for the cultivation unit operation of a production process for a whole cell vaccine against whooping cough, caused by the *Bordetella pertussis* bacterium. In this batch cultivation process the most critical parameters are the starting materials[1], because they are the only process input. In a batch process a small variation at the start can often result in a much larger variation at the end of the process. After the process has been started, only a few parameters can be controlled on line. For a typical batch cultivation process these parameters are usually pH, temperature and dissolved oxygen concentration. Because each of these parameters can influence bacterial growth and product formation, they classify as being critical to product quality and performance.

For a full PAT application, the failure limits of the critical parameters need to be known. This means that for each of the critical parameters the ranges have to be explored that still yield good product quality. Within these failure limits a safe range can be selected that is in principle not critical to product quality or process performance. In this way specifications can be defined that are based on actual process understanding.

Because most processes usually have several critical parameters that can interact with each other, the final step is to perform a series of experiments in which all critical parameters are varied simultaneously. Such a series of experiments can be designed using basic statistical principles. This approach is generally known as experimental design or Design of Experiments (DoE)[2]. By applying DoE the previously defined specifications can be tested for unexpected interaction with each other.

In this study DNA microarrays were used to detect changes in cellular physiology caused by a 90 minute period of oxygen limitation and to investigate whether any of the earlier reported marker genes are affected, implicating consequences for product quality. By using DNA microarrays in such a way general product and process understanding is increased as well as general understanding of cellular physiology, because the genes affected by the disturbance are revealed. At the same time the risk of a certain critical parameter for product quality and performance can be assessed. The use of DNA microarrays as a tool to increase process understanding has become more common in the recent past[3,4,5]. Previously we have demonstrated the suitability of DNA microarrays for the investigation of a bacterial cultivation processes[1]. We reported a set of genes that is important for the quality and performance of a whole cell vaccine against whooping cough disease. These marker genes are associated with virulence, because it is only the virulent bacterium against which a protective immune response can be raised[6]. Therefore the expression of virulence factors is crucial for vaccine performance, especially those that are on the outside of the bacterium or that are excreted [7].

It has been hypothesized[8,9] that the virulence state of members of the *Bordetella* genus is associated with survival outside the host and adaptation to unfavorable conditions on one hand and colonization and infection of the host on the other hand. *Bordetella pertussis* can switch between these states[10,11,12], which each have specific proteins that are involved in the mechanisms necessary either for survival or colonization. The proteins involved in colonization and infection (so called virulence factors) are the targets against which a protective immune response is raised in a whole cell vaccine against whooping cough. Therefore, changes in the expression of these virulence factors have implications for product quality[6,13]. For a strictly aerobic bacterium, oxygen limitation could be a condition that triggers the *survival* state. This means that the virulence factors are no longer expressed in favor of proteins involved in survival outside the host, resulting in an ineffective vaccine.

In this paper disturbance of the dissolved oxygen concentration during the cultivation of *Bordetella pertussis* is investigated for its impact on expected product quality using DNA microarrays. Because the *Bordetella* family consists of strictly aerobic micro-organisms [14], oxygen supply is crucial for bacterial survival. Contrary to what was expected before the experiment, oxygen limitation was found to be fully reversible on the gene expression level at the end of cultivation after having a large impact immediately after the disturbance. This finding has implications on how the risk of oxygen limitation on process and product performance should be assessed. It illustrates the importance of investigating a pharmaceutical manufacturing process, even for parameters that are considered well understood. True process understanding can only come from sound scientific research and careful scrutinization of all critical process parameters.

Materials and Methods

Bacterial cultivation

Seedlot and starter culture

Bordetella pertussis strain 509 (NVI, Bilthoven, The Netherlands) used in this study was collected in 1963 as a clinical isolate and was used as one of the two strains in the Dutch vaccination program until 2005. Cultivation was done using the chemically defined, filter sterilized THIJS medium containing lactate and glutamate as main carbon sources[15,16]. All starter cultures were inoculated from the same working seedlot as described earlier[1]. The working seedlot was prepared using freeze-dried *Bordetella pertussis* strain 509, obtained from the in-house culture collection. The culture was mixed with glycerol (10 % v/v) after growth to OD_{590}=1.0, divided in 10 ml aliquots and stored at -140°C. Shakeflask pre-cultures for the bioreactor experiments were inoculated at 5 % (v/v) from this seedlot, and cultured at 35°C to OD_{590} =1.0 using an orbital shaker at 200 rpm.

Bioreactor conditions and operation

Two fully instrumented bench-top bioreactors (Applikon, Schiedam, The Netherlands) with a total volume of 7 liters were used in this study. Bioreactors were fitted with 6-bladed Rushton turbines. Cultivations were performed at a working volume of 4L, using an inoculum of 5% v/v with OD_{590}=1.0. Temperature, pH and dissolved oxygen (DO) were controlled at 34 ºC , 7.2 and 30 %, respectively. A low-drift polarographic electrode (Applikon, Schiedam, The Netherlands) measured the dissolved oxygen in the medium, while pH and temperature were measured by a glass pH electrode (Mettler Toledo, Tiel, The Netherlands) and a Pt100 temperature sensor (Applikon, Schiedam, The Netherlands), respectively. All sensors were connected to the bioreactor control system ADI-1040 (Applikon, Schiedam, The Netherlands), which was operated using BCSV software (Compex, Ghent, Belgium).
Oxygen was supplied through the headspace only. Initially, oxygen demand was met by 50 rpm increments in stirrer speed from 450 rpm to 650 rpm. Further demand was met by increasing the oxygen fraction in the air flow with pure oxygen. Total gas flow was continuously kept constant at 1 L/min.
Seven batches were run, i.e. three control experiments (without oxygen disturbance) and four oxygen limited experiments, in which the pure oxygen supply was closed for a period of 90 minutes, while leaving the air supply intact at 1 L/min. The pure oxygen supply was closed when a certain biomass density was reached (OD_{590}=0.4 ± 0.04). At this point the stirrer had reached maximum speed and a fraction of pure oxygen was mixed with air and fed into the bioreactor headspace to maintain DO setpoint. By closing the pure oxygen supply, the bioreactor remained aerated, but oxygen limited. This resulted in a sharp decrease in

dissolved oxygen concentration to 0% (Figure 1). The 90 minute period of oxygen limitation was considered to start when DO < 1%. When pure oxygen supply was restored, DO increased sharply and returned back to its original setpoint within a few minutes. This method of oxygen limitation was chosen because it allowed good control of the length of oxygen limitation, resulting in reproducible experiments and because it is the only way to isolate the effect of oxygen limitation from secondary effects. Complete cut-off of the gas-flow would also eliminate carbon dioxide ventilation and cause unwanted side effects. Venting with pure nitrogen could result in a complete arrest of cellular function and even cell death and lysis because *Bordetella pertussis* is an obligate aerobic organism. This would make microarray analysis impossible and is not a likely situation to occur during manufacturing.

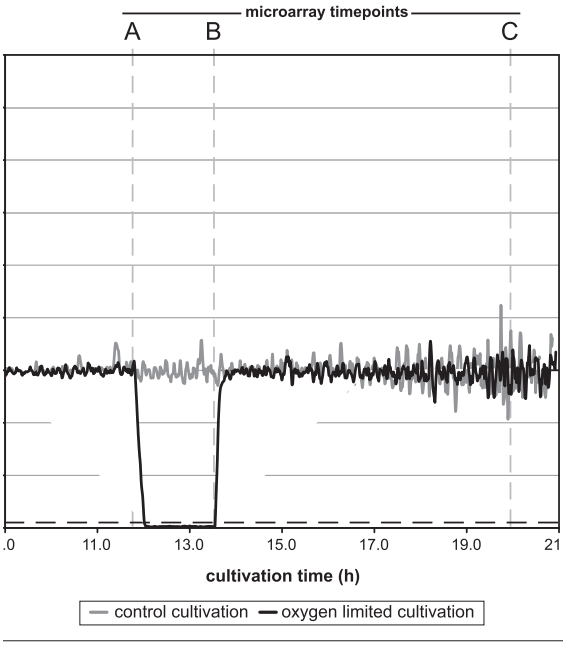


Figure 1. Dissolved oxygen concentration profiles of an oxygen limited (dark line) culture and a control culture (light line) between 9 h and 21 h of culture time. The sudden drop in the dark profile is caused by the closure of the pure oxygen supply into the headspace, while the air supply remained intact. After 90 minutes the pure oxygen supply was restored and the dissolved oxygen concentration quickly returned to set-point (30%). The dashed line indicates 1% dissolved oxygen concentration, which is the start of the 90 minutes oxygen depletion. The letters A, B and C indicate the samples taken immediately before oxygen limitation, immediately after oxygen limitation and at the end of the batch, respectively.

Sampling

Samples were taken from the reactor at regular intervals. Biomass concentration was measured offline by photospectroscopy at 590 nm using a Vitalab 10 Photospectrometer (Vital Scientific, Dieren, The Netherlands) at 590 nm. L-lactate concentrations were determined *at line* using an YSI 2700 analyzer (Yellow Springs Instruments, Yellow Springs, USA). These data was used for end-point determination ([lactate] = 2–3 mM). Supernatant of samples was stored and later analyzed by ^{1}H NMR (JEOL JNM-ECP 400FT, JEOL, Tokyo, Japan) for other nutrient concentrations and also to check for any detectable disturbances in the

metabolism such as the excretion of waste metabolites caused by the oxygen limitation. Just before interrupting (A) and just before restoring (B) the oxygen flow, samples for mRNA analysis were taken. At the end of the batch samples for both protein and mRNA analysis were taken (C). The batch was stopped when the residual lactate concentrations was between 2 and 3 mM. Samples in the control cultures were taken at or near the same culture times as the oxygen limited cultures for samples A and B and at the same lactate concentrations for sample C (2-3 mM).

RNA Analysis

MicroArray design

Samples were analyzed using a full genome *Bordetella pertussis* microarray. Based on the complete genome sequence of *Bordetella pertussis* Tohama I[17], a set of 3582 70-mer oligonucleotides was developed at Operon (Cologne Germany), covering 93% of all open reading frames. Oligonucleotide pellets were dissolved in 50% DMSO (v/v in water) to a concentration of 20μM and spotted in triplicate on UltraGAPS II coated slides (Corning, New York U.S.A.), together with 420 control spots, using the Omnigrid 100 microarray spotter (GeneMachines, San Carlos U.S.A.). Prior to use, slides were UV cross-linked, treated with sodium borohydride and pre-hybridized according to the manufacturer's protocol (Corning, New York, USA).

RNA isolation

For fixation of the RNA expression profile, 1 volume bacterial culture was mixed with 2 volumes RNAse retarding solution[18,19,20]. For each microarray sample 2.5 ml bacterial suspension at OD_{590} = 1.0 was used. Samples at other optical densities were adjusted accordingly, therefore an equal amount of cells was used for each RNA isolation step. The samples were concentrated by centrifugation and treated with Tris-EDTA buffer, containing 0.5mg/ml lysozyme (Sigma-Aldrich, Zwijndrecht, The Netherlands) for 3 minutes. Total RNA was extracted with the SV Total RNA isolation system (Promega Benelux, Leiden, The Netherlands) according to manufacturer's protocol. Nucleic-acid concentration was adjusted by precipitation and spectral analysis was used to determine final nucleic-acid concentration and purity. RNA integrity was confirmed with the Bioanalyzer RNA6000 Nano assay (Agilent Technologies, Amstelveen, The Netherlands), according to the manufacturer's protocol.

Labeling and Hybridization

Total RNA from all experimental samples was reverse transcribed to cDNA and labeled with Cy3/Cy5 dyes using the Chipshot Indirect Labelling Kit (Promega Benelux, Leiden, The Netherlands) according to manufacturer's protocol, with one deviation: only 2µl random nonamer primer and no oligo-dT primer was used per reaction to reverse transcribe 50µg total

RNA. The labeled and purified cDNA samples were pooled in Cy3/Cy5 pairs and volumes were adjusted to 25µl by evaporation under low pressure. Hybridization buffer (25µl) was added, to a final concentration of 25% formamide, 5x SSC and 0.1% SDS. Samples were applied on the microarray slides and incubated in a hybridization chamber (GeneMachines, San Carlos, U.S.A.) for 16-20 hours at 42°C in the dark.

Data acquisition and analysis

The microarrays were scanned with a ScanArray Express microarray scanner (Perkin Elmer, Groningen, The Netherlands) and median fluorescence intensities were quantified for each spot using ArrayVision software (Imaging Research, Roosendaal, The Netherlands). The expression data were natural-log transformed, quantile normalized, and values of replicate spots were averaged. These data processing steps were done with the free statistical software R (WU Wien, Austria), using an in-house developed script.

Differential gene expression levels were obtained through comparison with a common reference sample, containing equal amounts of RNA for each sample. P-values were calculated using one-way ANOVA statistical analysis and adjusted for multiple testing by calculating the False Discovery Rate (FDR). Fold Ratio (FR) values were expressed as the natural log of the normalized signal ratio. Observed differential gene expression was considered biologically significant when Fold Ratio (FR) > 1.5 and False Discovery Rate (FDR) < 0.10. Additional data analysis by principal component analysis (PCA) was done using GeneMaths software (Applied Maths, Sint-Martens-Latem, Belgium).

Gene expression shifts between PCA groups (Figure 3) were assessed using Eucledian difference in ln-transformed gene expression values to calculate the Mahalanobis distances over the set of significantly regulated genes. This corresponds to the distance in the PCA plot. P-values were calculated by using these distances in a Hotelling T-test.

Results

Culture characteristics

All seven cultures showed comparable overall growth characteristics and nutrient consumption rates. The control cultures seemed to have a slightly higher biomass yield and associated lactate consumption, but they also had a slightly higher lactate concentration at the start of the batch. Overall, no significant differences were found between both groups in biomass yield from the start of oxygen limitation until the end of the batch ($p<0.21$), i.e. between sample points A and C (Figure 1). The NMR nutrient analysis did not reveal any change in excreted (waste) metabolites or other low molecular weight compounds after oxygen limitation.

Microarray results

The complete dataset from the microarray experiment can be found in the supplementary data file (Table A). Data from samples taken before oxygen limitation (A), immediately after oxygen limitation (B), and at the end of the culture (C) and the samples taken from the control cultures at the corresponding time points were averaged between replicates. Genes that showed a significant difference between any of the samples were considered a hit and are shown in Figure 2 (see on line supplementary datafile Table A for p-values). The 104 genes identified as hits cluster in 7 groups. Groups I and II consist of 34 genes that were respectively only up- or down-regulated immediately after the oxygen limitation period. Groups III and IV consist of 41 genes that were respectively up- or down-regulated at the end of the cultivation (end of batch effect), but not by the oxygen limitation. Groups V and VI consist of 24 genes that were respectively up- or down-regulated immediately after oxygen limitation and (to varying extents) at the end of the batch. Group VII consists of 5 genes that are variably regulated. The genes in the groups I, II, V and VI are the genes that are directly affected by oxygen limitation, i.e. these are the genes that showed a significant hit between the oxygen limited sample B and the control sample B.

Genes that have consecutive gene numbers (i.e. are located next to each other on the genome) and that were regulated similarly were marked as potential operons, i.e. genes that were regulated together, usually to perform a specific cellular task. Table B in the on line supplementary datafile (journal homepage) shows an overview of these potential operons and their possible functions. Most prominently, genes involved in mRNA and protein synthesis and oxidative phosphorylation are regulated in an operon-like manner as a result of the oxygen limitation (groups I, II and VI). Towards the end of the batch, genes involved in nutrient transport and metabolism show an effect, most prominently sulphate transport (group III).

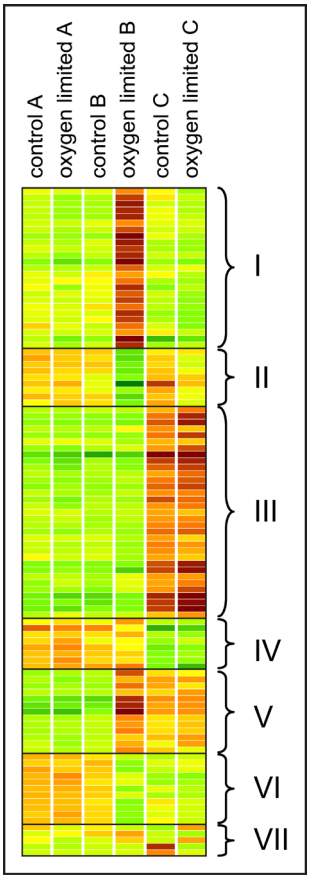


Figure 2. Relative expression levels of all genes that showed a significant (FR>1.5, FDR<0.1) result in any of the groups. Sample points A, B and C indicate samples taken before oxygen limitation, immediately after oxygen limitation and at the end of the cultivation. The control samples were taken at the same culture times as the oxygen limited samples for A and B and at the same lactate concentration for sample C. The color scale runs from dark green to dark red through yellow, where dark green indicates strong down-regulation and dark red indicates strong up-regulation.

Covariation between samples can be visualized with principal component analysis (PCA). Figure 3 shows how the hits from the respective samples are distributed when the first principal component is plotted against the second principal component. It is clear that sample B (right after oxygen limitation) of the experimental group clusters separately from the A samples (before oxygen limitation) and the samples A and B from the control group ($p<0.0023$). Surprisingly, the samples C (end of batch) from the control and the oxygen limited group all cluster together, indicating no difference between both groups at the end of the batch.

In order to assess the impact on product quality, the found hits were screened for the presence of any of the earlier reported 56 virulence core regulon genes (virulence activated genes or vag's) that can serve as a marker for product quality[1]. Two virulence genes, bcrH1 and bcrH2, were found as significant hit in any one of the samples (see supplementary data table A). These are both part of the Type III secretion system operon[21], that has recently been identified

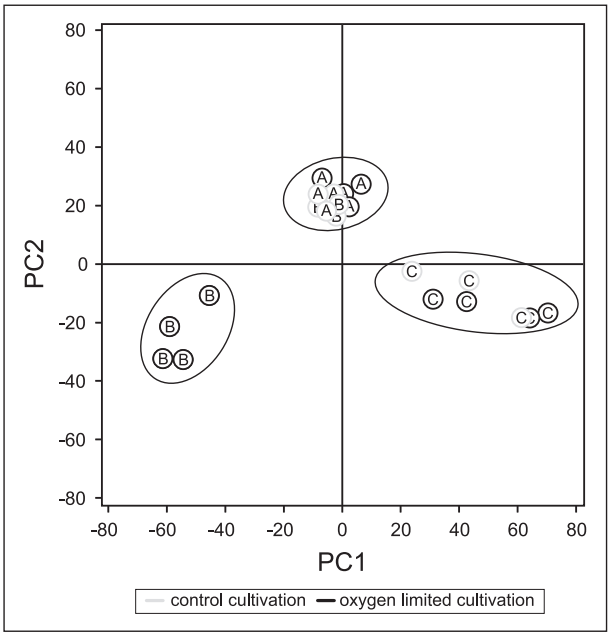


Figure 3. Principal component analysis of the hits found with DNA microarrays. This plot shows that the samples before oxygen limitation (dark and light circled A) cluster in the center, together with the control samples at the time point immediately after oxygen limitation (light circled B). On the left hand side the samples immediate after oxygen limitation (dark circled B) of the oxygen limited cultivations are clustered. This cluster is significantly different from the center cluster ($p<0.0023$). All samples taken at the end of cultivation (dark and light circled C) cluster together at the right-hand side. This cluster is also significantly different from the center cluster ($p<0.013$). This indicates that the differences between the control cultivations and oxygen limited cultivations observed at time point B are reversed at the end of the cultivation, but that there is a shift in gene expression towards the end of cultivation that is identical for both control and oxygen limited cultivations.

to be involved in the excretion of virulence factors in *Bordetella pertussis* [22]. However, neither showed an effect in the oxygen limited sample B as compared to the control sample B, but these genes were only identified as hit between samples A and C, indicating a change of expression towards the end of cultivation.

Discussion

The experiments described in this paper were performed with the aim to find the effects on the end product caused by an oxygen limitation during cultivation in order to assess the criticality of oxygen limitation for product quality. Surprisingly, at the end of the cultivation no difference could be found between the oxygen limited cultures and the control cultures using DNA microarrays, despite the fact that immediately after the oxygen limitation a clear effect was observed on the expression levels of numerous genes (Figure 2 and on line supplementary data). This means that oxygen limitation has a fully reversible effect on the physiology of *Bordetella pertussis* and is less critical for product quality than previously thought.

The data from the microarrays prove highly valuable, because they give information about every single gene of the organism, which allows the observation of changes in pathways or cellular functions that with any other technique would remain undiscovered. Also, a process disturbance that has such a high impact at the moment it occurs and is undetectable with microarrays at the end of the process can be reliably labeled as reversible and istherefore less critical than was believed before. In this case, the genes involved in oxygen limitation did not show a correlation with the genes of the virulence core regulon. From this the criticality of oxygen limitation on product quality can be assessed as limited and reversible.

However, two virulence associated genes were identified as a hit between samples A and C. Although two genes are only a small fraction of the reported 56 virulence genes[1], this is still a worrying finding, because it could indicate that towards the end of a batch the virulence of *Bordetella pertussis* is affected. This notion is consistent with the idea that the *bvg* / virulence system of the *Bordetella* family is intrinsically evolved to deal with survival outside a host in a low-nutrient environment[8,9]. At the end of a cultivation batch the bacteria experience a low-nutrient environment as well. Although the system is not fully functional in *Bordetella pertussis*, it does still react to the same environmental stimuli as the functional system in *Bordetella bronchoseptica*[8] and could therefore very well play a critical role in product quality at the end of a production batch.

In conclusion, the results in this paper take us one step further in the investigation of the critical process parameters of the cultivation process of *Bordetella pertussis*. This work shows that even a parameter that is generally accepted as being critical in a bioprocess appears to have only transient effects on process performance. The specifications for the production of a vaccine against whooping cough might be adjusted accordingly, although the analysis presented in this paper do not directly link to the clinical performance of the product. However as mentioned earlier, a link between the virulence state and product

performance has already been established[6]. The regulatory flexibility necessary to change and adjust specifications based on scientific evidence is provided in the FDA's PAT guidance document (http://www.fda.gov/cder/guidance/6419fnl.pdf).

In order to fully comply with the PAT framework, the interaction between oxygen limitation and other critical process parameters needs to be evaluated further. Only then can a process disturbance be accurately assessed for its risk on product quality. Finally these data and possibly additional online data need to be combined in a readily (i.e. on line) available database that can be used as cross reference during new manufacturing runs. This allows for an on line check on the performance of the running process in respect the already available process understanding, allowing the assurance of process performance and product quality in real time.

Acknowledgements

The authors gratefully acknowledge Dr. J. Pennings for help with the DNA microarray dataprocessing and Dr. B. Zomer for NMR nutrient analyis. This work was partly funded by Senter Novem, The Netherlands, grant number TSGE3067.

References

1. Streefland M, van de Waterbeemd B, Happe H, van der Pol LA, Beuvery EC, Tramper J, Martens DE. 2007. PAT for vaccines: The first stage of PAT implementation for development of a well-defined whole-cell vaccine against whooping cough disease. Vaccine 25(16):2994-3000.
2. Montgomery DC. 1999. Experimental Design for Product and Process Design and Development. The Statistician 48(2):159-177.
3. Hirasawa T, Nakakura Y, Yoshikawa K, Ashitani K, Nagahisa K, Furusawa C, Katakura Y, Shimizu H, Shioya S. 2006. Comparative analysis of transcriptional responses to saline stress in the laboratory and brewing strains of Saccharomyces cerevisiae with DNA microarray. Appl Microbiol Biotechnol 70(3):346-57.
4. Khodursky AB, Peter BJ, Cozzarelli NR, Botstein D, Brown PO, Yanofsky C. 2000. DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in Escherichia coli. Proc Natl Acad Sci U S A 97(22):12170-5.
5. Seth G, Charaniya S, Wlaschin KF, Hu WS. 2007. In pursuit of a super producer-alternative paths to high producing recombinant mammalian cells. Curr Opin Biotechnol 18(6):557-64.
6. Bellalou J, Relyveld EH. 1984. Studies on culture conditions of Bordetella pertussis and relationship to immunogenicity of vaccines. Ann.Microbiol.(Paris) 135B(1):101-110.
7. Mills KH. 2001. Immunity to Bordetella pertussis. Microbes.Infect. 3(8):655-677.
8. Cummings CA, Bootsma HJ, Relman DA, Miller JF. 2006. Species- and strain-specific control of a complex, flexible regulon by Bordetella BvgAS. J.Bacteriol. 188(5):1775-1785.
9. Nakamura MM, Liew SY, Cummings CA, Brinig MM, Dieterich C, Relman DA. 2006. Growth phase- and nutrient limitation-associated transcript abundance regulation in Bordetella pertussis. Infect.Immun. 74(10):5537-5548.
10. Cotter PA, Jones AM. 2003. Phosphorelay control of virulence gene expression in Bordetella. Trends Microbiol. 11(8):367-373.
11. Scarlato V, Prugnola A, Arico B, Rappuoli R. 1990. Positive transcriptional feedback at the bvg locus controls expression of virulence factors in Bordetella pertussis. Proc.Natl.Acad.Sci.U.S.A 87(17):6753-6757.
12. Smith AM, Guzman CA, Walker MJ. 2001. The virulence factors of Bordetella pertussis: a matter of control. FEMS Microbiol.Rev. 25(3):309-333.
13. Hamstra HJ, Kuipers B, Schijf-Evers D, Loggen HG, Poolman JT. 1995. The purification and protective capacity of Bordetella pertussis outer membrane proteins. Vaccine 13(8):747-752.
14. Goodnow RA. 1980. Biology of Bordetella bronchiseptica. Microbiol Rev 44(4):722-38.
15. Thalen M, van den IJssel J, Jiskoot W, Zomer B, Roholl P, de Gooijer C, Beuvery C, Tramper J. 1999. Rational medium design for Bordetella pertussis: basic metabolism. J.Biotechnol. 75(2-3):147-159.
16. Thalen M, Venema M, van den IJssel J, Berwald L, Beuvery C, Martens D, Tramper J. 2006. Effect of relevant culture parameters on Pertussis Toxin expression by Bordetella pertussis. Biologicals 34(3):213-220.
17. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MT, Churcher CM, Bentley SD, Mungall KL and others. 2003. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. Nat.Genet. 35(1):32-40.
18. Kasahara T, Miyazaki T, Nitta H, Ono A, Miyagishima T, Nagao T, Urushidani T. 2006. Evaluation of methods for duration of preservation of RNA quality in rat liver used for transcriptome analysis. J Toxicol Sci 31(5):509-19.
19. Mutter, GL, Zahrieh D, Liu C, Neuberg D, Finkelstein D, Baker HE, Warrington JA. 2004. Comparison of frozen and RNALater solid tissue storage methods for use in RNA expression microarrays. BMC Genomics 5(1):88.

20. Ramalho AS, Beck S, Farinha CM, Clarke LA, Heda GD, Steiner B, Sanz J, Gallati S, Amaral MD, Harris A and others. 2004. Methods for RNA extraction, cDNA preparation and analysis of CFTR transcripts. J Cyst Fibros 3 Suppl 2:11-5.
21. Mattoo S, Yuk MH, Huang LL, Miller JF. 2004. Regulation of type III secretion in Bordetella. Mol.Microbiol. 52(4):1201-1214.
22. Fennelly NK, Sisti F, Higgins SC, Ross PJ, van der Heide H, Mooi FR, Boyd A, Mills KH. 2008. Bordetella pertussis expresses a functional type III secretion system that subverts protective innate and adaptive immune responses. Infect Immun.

CHAPTER 4

Gene-expression based quality scores indicate optimal harvest point in *Bordetella pertussis* cultivation for bacterial vaccine production

Mathieu Streefland[1]*, Bas van de Waterbeemd[1]*, Jeroen Pennings[2], Leo van der Pol[1], Coen Beuvery[3], Johannes Tramper[4], Dirk Martens[4]

*both authors contributed equally

[1]Netherlands Vaccine Institute (NVI), Bilthoven, The Netherlands.
[2] National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands
[3]PAT Consultancy, Vianen, The Netherlands.
[4]Wageningen University, Wageningen, The Netherlands.

Accepted for publication in Biotechnology and Bioengineering

Supplementary data can be found with the on line published version of this chapter at the journal's home page

Abstract

Introduction

The evolution of vaccine product quality during batch cultivation of *Bordetella pertussis*, the causative agent of whooping cough, was investigated with the goal to determine the optimal harvest point. The process was explored by measuring mRNA expression at frequent intervals during cultivation. Because the virulence genes that are important for quality are already known for this product, changes in expression levels of these genes are proposed to be indicative for product quality.

Results

A quantitative product quality score is calculated, based on the expression levels of the genes that are critical for product quality. Product quality scores were maximal throughout the logarithmic growth phase, but dropped significantly at the end of the logarithmic phase and the start of the stationary phase, caused by nutrient depletion. The optimal harvest point for this batch process was determined, based on rational science-based criteria.

Conclusions

By measuring the changes in expression levels of genes that are critical for product quality during cultivation, the optimal harvest point of *Bordetella pertussis* cultivation has been determined. The product quality score, based on the expression of these genes, allows comparison of expected product quality between culture samples or between batches. This showed that the decreasing lactate and glutamate concentrations towards the end of the batch are critical for product quality. On line measurement of these nutrients allows the cultivation process to be harvested at or near the optimal harvest point, increasing process robustness and consistency.

Introduction

Batch cultivation processes are common in biopharmaceutical manufacturing. GMP guidelines and other regulations have delayed the evolution to more continuous processing, which is much more common in other process industries (i.e. petrochemical industry). Batch cultivations have several drawbacks compared to continuous processing. First of all, the nutrients and other substrates all need to be present at the start and their starting concentration defines the maximal yield at the end of the batch. Moreover, the quality of the starting materials is crucial for batch performance. Also much time is lost between batches (downtime) and batch processing is relatively labor intensive, because of cleaning and handling. Of course the yield can be increased by adopting a fed-batch strategy, but the downstream processing usually still involves batch-wise unit operations, which makes continuous harvesting impossible.

Besides the importance of the starting conditions for a batch process, the determination of the end-point might be more important for product quality. At the end of a bacterial batch cultivation the conditions permitting growth have ceased to exist and the biomass will start to die off. This death phase is usually preceded by a stationary phase in which enough nutrients are available to sustain basal maintenance metabolism, but not enough to sustain growth. These phases are usually marked by profound changes (i.e. in metabolic and other pathways) inside the cell[1,2,3]. For a biopharmaceutical product, the final phase of the batch may be critical to the quality of the product. At the end of a bacterial cultivation cells usually lyse, which is accompanied by the release of unwanted proteins, DNA or enzymes (proteases) that need to be separated from the active pharmaceutical ingredient (API) or can degredate the product. In the case of a whole cell product, changes in cellular morphology or outer membrane composition can be critical to product performance.

The optimal harvest point of a bacterial batch cultivation should result in the highest possible yield of product at the desired level of product quality. However, both could very well be in direct competition towards the end of the batch. Product quality even could be a variable throughout the batch or could be a constant for most of the batch and only change towards the end of the batch. In order to determine the optimal harvest point and to increase the understanding of the evolution of product quality during a batch cultivation, the entire batch profile needs to be investigated by frequent sampling and analysis of product quality and process performance. This paper describes the investigation of the batch profile of the cultivation process step for the manufacturing of a whole cell vaccine against whooping cough disease, a severe childhood disease caused by the *Bordetella pertussis* bacterium

The manufacturing process of this whole cell vaccine only consists of a cultivation process step followed by heat inactivation and concentration to yield the bulk product. For such a product it is especially critical to investigate the cultivation process step, because this is where the actual product quality is determined (i.e. the potency of the vaccine) and the limited downstream processing can change or add little to product quality. The starting conditions and materials that yield a robust batch process for *Bordetella pertussis* cultivation have been investigated earlier[4].

The condition and composition of the biomass, more specifically the outer membrane, is especially critical for a whole cell vaccine, because it contains the antigens that induce a protective immune response after vaccination. Most important antigenic outer membrane proteins are, however, known to change with the changing process conditions associated with batch processing. Nakamura *et al.*[5] have reported that after the nutrients had been completely consumed, substantial changes in gene expression patterns were observed in *Bordetella pertussis*, including many genes of the so called virulence core regulon [6,4]. The genes of this regulon encode virulence proteins that are involved in host invasion and further infection and many of them are known to induce or to play a role in the induction of a protective immune response after vaccination[7]. Their importance is confirmed by the fact that immunization studies with non-virulent *Bordetella pertussis* bacteria showed the loss of a protective immune response[8]. Purified forms of these proteins are commonly used in varying combinations in a-cellular vaccines against whooping cough disease.

Previously[4], we have demonstrated that microarrays are a suitable tool to identify and quantitatively measure the relative expression levels of the virulence core regulon genes. This allows discrimination between favorable and unfavorable process conditions by assessing the impact of different process conditions on the relative expression levels of these marker genes. In this paper this principle is used to determine the optimal harvest point and to investigate how the continuously changing conditions of batch processing are reflected in the gene expression profile. mRNA expression profiles are measured throughout logarithmic growth and plateau phases to increase the understanding of how the bacterium adapts to the continuously changing environment of a batch process. Moreover, by measuring the expression of the core regulon genes, the impact of these changes on expected product quality can be determined.

The correlation between nutrient availability and expected product quality allows prediction and control of the optimal batch end point. At line measurements of nutrient concentrations allow harvest at the highest yield without endangering product quality. In fact, observed in the light of the recent Process Analytical Technology (PAT) initiative[9], we now have a

parameter that is critical for product quality and that can be measured during cultivation. The at line measurements can be replaced by online monitoring of nutrient concentration with spectroscopic techniques, such as near infrared spectroscopy. On line monitoring and control of critical process and product attributes with the aim of ensuring final product quality is what defines PAT according to the FDA's guidance on Process Analytical Technology[9].

This paper demonstrates that the overall expression pattern of *Bordetella pertussis* mRNA is relatively constant during cultivation, but towards the end of the logarithmic phase, the decreasing availability of nutrients suddenly affects a large number of genes after the key nutrients lactate and glutamate drop below a minimal threshold concentration. Among these genes are many of the genes involved in virulence, which is an indication that product quality is at risk when the bacteria are harvested too late.

The level of bacterial virulence of culture samples is scored on a scale between optimal virulence and complete non-virulence using principal component analysis. This allows us to give a virulence-dependent quality score to cultivation samples and especially harvest samples. In this way product quality at different sample points can be compared without having to use animal potency tests for research activities. Ultimately such a score can be used to compare different cultivation end-products with each other or with a *golden standard* reference. In the mean time such a "quality scale" can serve as a powerful research tool during process development.

Materials and Methods

Bacterial strain, media and growth conditions

The *Bordetella pertussis* 509 strain (Netherlands Vaccine Institute, Bilthoven, The Netherlands) used in this study was collected in 1963 as a clinical isolate and used as vaccine strain in the combined diphtheria, pertussis, tetanus and polio (DPTPolio) vaccine until 2005. All bacterial cultures were grown in chemically defined THIJS medium[10,11], which consists of basic medium and supplement. Supplement (1%, v/v) was added to the basic medium shortly before inoculation. The pH after addition of supplement is 7.1. The preculture was inoculated with 10 ml *Bordetella pertussis* 509 working seedlot (cells at OD_{590} = 1.00 ± 0.05 frozen with 10% glycerol, stored at -140° C) and grown in 500 ml shake flasks containing 200 ml medium at 35° C, on an orbital shaker at 200 RPM. When the starting culture reached OD_{590} = 1.00 ± 0.05, it was used to inoculate a bioreactor. Start density was OD_{590} = 0.050 (± 0.005) for all cultivations.

Bioreactor cultivations

All cultivations were carried out in a fully instrumented 7 L in situ sterilizable bench-top bioreactor with six-bladed Rushton turbine (Applikon, Schiedam, The Netherlands). After the bioreactor was filled with 4 L THIJS medium and reached set point temperature and 100% dissolved oxygen (DO), it was inoculated with 200 ml preculture with an optical density at 590nm (OD_{590}) of 1.00 ± 0.05. Temperature, DO, pH and stirrer speed were controlled at 35 ± 0.1° C, 30 ± 2%, 7.2 ± 0.1 and 450-650 rpm, respectively. A low-drift polarographic electrode (Applikon, Schiedam, The Netherlands) measured the dissolved oxygen concentration, while pH and temperature were measured by a glass pH electrode (Mettler Toledo, Tiel, The Netherlands) and a Pt100 temperature sensor (Applikon, Schiedam, The Netherlands), respectively. All sensors were connected to the ADI-1040 bioreactor control system (Applikon, Schiedam, The Netherlands), which was operated using BCSV software (Compex, Ghent, Belgium). DO was first controlled with increments of 50 rpm in stirrer speed until the 650 rpm maximum was reached. After this set point was reached, DO was controlled by increasing the fraction of oxygen in the headspace. Total gas flow was kept constant at 1.0 L/min. Samples were taken at time points indicated in Figure 1 for OD_{590} and nutrient concentration measurements. Samples were sterile filtered (0.22 μm) and stored at -20° C for nutrient analysis. Growth and nutrient consumption data were fitted using Tablecurve 2D software (Systat Software, San Carlos, CA, U.S.A.).

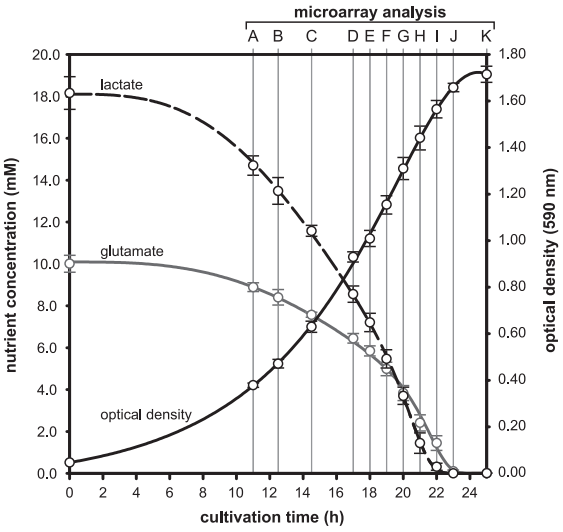


Figure 1. Fitted growth and nutrient consumption curves of the 4 identically operated bioreactor cultivations. Right Y-axis represents growth (optical density at 590 nm) and left Y-axis represents concentrations of the primary carbon and nitrogen sources in the growth medium (mM). Samples were taken at 11 time points (A to K), indicated by grey vertical lines, that correspond with samples for microarray analysis. Error bars indicate standard error (SE).

Nutrient concentration analysis

Lactate and glutamate concentrations were determined with 1H-NMR using a JEOL JNM ECP 400 spectrometer operating at 400MHz (JEOL, Tokyo, Japan) equipped with a JEOL Stacman auto sampler for 16 samples. Supernatants were analyzed by adding 0.1 ml of D_2O containing 3-(trimethylsilyl)[D4]proprionic acid sodium salt (TMSP, 0.167 mM) to a 0.9 ml sample. The water signal was suppressed by irradiating the signal with standard NMR software. The spectra were referenced using the TMSP signal at 0 ppm. Lactate and glutamate concentrations were quantified by integration of the relevant signals. NMR was also used to check samples for any waste metabolites.

Microarray design

Samples were analyzed using a full genome *Bordetella pertussis* microarray. Based on the complete genome sequence of the *Bordetella pertussis* Tohama I strain[12], a set of 3582 70-mer oligo nucleotides was developed at Operon (Cologne Germany), covering 93% of all open reading frames. Oligo nucleotide pellets were dissolved in 50% DMSO (v/v in water) to a concentration of 20 μM and spotted in triplicate on UltraGAPS II coated slides (Corning, New York, NY, U.S.A.), together with 420 control spots, using the Omnigrid 100 microarray spotter (GeneMachines, San Carlos, CA, U.S.A.).

RNA isolation

For fixation of the RNA expression profile, 1 volume bacterial culture was mixed with 2 volumes *RNase retarding solution*[13,14,15]. For each microarray sample 2.5 ml cultivation at

OD_{590} = 1.0 was used. Samples at other optical densities were adjusted accordingly, so that an equal amount of cells was used for each sample. The samples were concentrated by centrifugation and treated with Tris-EDTA buffer, containing 0.5 mg/ml lysozyme (Sigma-Aldrich, Zwijndrecht, The Netherlands) for 3 minutes. Total RNA was extracted with the SV total RNA isolation system (Promega Benelux, Leiden, The Netherlands) according to manufacturer's protocol. Nucleic acid concentration was adjusted by precipitation and spectral analysis was used to determine final nucleic acid concentration and purity. RNA integrity was confirmed with the Bioanalyzer RNA6000 Nano assay (Agilent Technologies, Amstelveen, The Netherlands), according to the manufacturer's protocol.

cDNA labeling and hybridization reactions

Total RNA from all experimental samples was reverse transcribed to cDNA and labeled with Cy3/Cy5 dyes using the Chipshot Indirect Labeling kit (Promega Benelux, Leiden, The Netherlands) according to manufacturer's protocol, with one deviation: 2 µl random nonamer primer and no oligo-dT primer was used per reaction to reverse transcribe the total RNA. Experimental samples (Cy5) were pooled with a common reference sample (Cy3), containing equal amounts of RNA from all experimental samples in the experiment. Volumes of the combined cDNA samples were adjusted to 25 µl and an equal amount of hybridization buffer was added, to a final concentration of 25% formamide, 5x SSC and 0.1% SDS. Samples were applied to the microarray slides and placed in a hybridization chamber (GeneMachines, San Carlos, CA, U.S.A.) for 16-20 hours at 42° C in the dark.

Data acquisition and analysis

The microarrays were scanned with a ScanArray Express microarray scanner (Perkin Elmer, Groningen, The Netherlands) and median fluorescence intensities were quantified for each spot using ArrayVision software (Imaging Research, Roosendaal, The Netherlands). The expression data were natural-log transformed, quantile normalized, corrected for the common reference dye signal and values of replicate spots were averaged. These data processing steps were done with the free statistical software R (WU Wien, Austria), using an in-house developed script. P-values were calculated using one-way ANOVA statistical analysis and adjusted for multiple testing by calculating the False Discovery Rate (FDR). Maximal Fold Ratio (FR) values were expressed as the natural log of the maximal normalized signal difference between any of the time points. A FDR of 1.0% was used to select genes whose gene expression showed a statistically significant difference. To further select for biologically relevant effects, only statistically significant genes with a maximal FR above 0.69 (or 2.00 without natural-log transformation) were included in the final analysis.

Principal Component Analysis (PCA)[16,17] and arrangement of differentially expressed genes in expression group clusters was performed with Genemaths 2.01 software (Applied Maths,

Sint-Martens-Latem, Belgium). Overall gene expression differences between time point groups were assessed by calculating the Euclidian distance between natural log transformed gene expression values over the entire set of genes (this corresponds to the distance in a PCA-plot). P-values were calculated by using these distances in a Hotelling T-test. Functional categorization of genes was done using the pathway annotations provided by the KEGG database. Pathway abundance in expression groups was estimated by calculating the binomial distribution probability. Pathways with P-values below 0.05 and also a representation of at least 4 genes in any of the expression groups were considered significantly more abundant than expected by chance.

To calculate product quality scores, only genes from the *virulence core regulon*[4] were considered. The results from this study were combined with data from a shake flask study, aimed to determine the product quality extremes. This was done by comparing the virulence core regulon gene expression pattern during mid-logarithmic growth of 4 shake flasks with normal growth conditions to 8 shake flasks with virulence repressing growth conditions (addition of $MgSO_4$ or niacin to the growth medium[4]). Natural log transformed gene expression values were adjusted with the median of all samples that were expected to give high product quality. PCA analysis was done on the resulting values and the scores of the first principal component (PC1) were used together with the expression data to fit a linear model containing a weight factor for each virulence core regulon gene. Product quality scores were calculated on an arbitrary 0 to 10 scale, with a linear relation to the weighted PC1 scores.

Results

Bacterial growth and nutrient consumption

Quadruplicate bioreactor cultivations were sampled at 11 time points, A to K (Figure 1), for growth (OD_{590}), lactate concentration (carbon source), glutamate concentration (both nitrogen source and carbon source) and microarray analysis (see paragraph 3.2). Logarithmic growth was observed from time points A to H. The culture becomes stationary during time points I, J and K. This growth arrest coincides with lactate depletion at around time point I, followed by glutamate depletion at around time point J (Figure 1). Variance on growth and nutrient concentrations was low throughout the experiment, as indicated by the error bars in Figure 1.

Microarray gene expression profiles

The RNA samples from time points A to K (Figure 1) were analyzed using full genome *Bordetella pertussis* DNA-microarrays[12,4]. The overall expression profiles of samples at identical time points were compared and showed high correlation for time points A to H ($R^2 = 0.990 \pm 0.003$) and good correlation for time points I to K ($R^2 = 0.962 \pm 0.030$). Principal Component Analysis[16,17] (PCA; Figure 2) was then used to identify the main sources of variance in the dataset. The first principal component (PC1) has a 67.9% contribution to total gene expression variance. The variance explained by PC1 reveals a significant difference between the samples at time points K when compared to samples at A to H ($p < 0.000054$), while time points I and J reveal a transitional phase between these two extremes. This PC1-shift coincides with the lactate and glutamate shortage and growth arrest shown in Figure 1. PC2 has only 8.0% contribution and explains the variance within group A to H by dividing it in 2 subgroups: time points A to C and D to H ($p < 0.151$; not significant). The other principal components had minor contributions to total gene expression variance and produced no significant shifts between time point groups.

The gene expression data was used to select a total of 705 genes that were significantly regulated at any of the 11 time points. These genes were clustered in time point-specific patterns, which resulted in 4 distinctive expression groups (figure 3A). Group I consists of 290 genes that are up regulated after time point H while genes from group II (355 genes) are down regulated after this point. Group III contains 52 genes that are up regulated during time points D to H. These genes are partly annotated to the sulfur metabolism pathway, sulfate and amino acid-related ABC transporters and to the *cys* and *met* operons, which are both involved in the biosynthesis of the sulfur containing amino acids[18]. Group IV contains the remaining 8 genes, with varying expression patterns. (The supplementary data file (table A1 to A4) contains a summary of the 4 expression groups, with gene ID's, annotations, average expression values per time point and statistical data.)

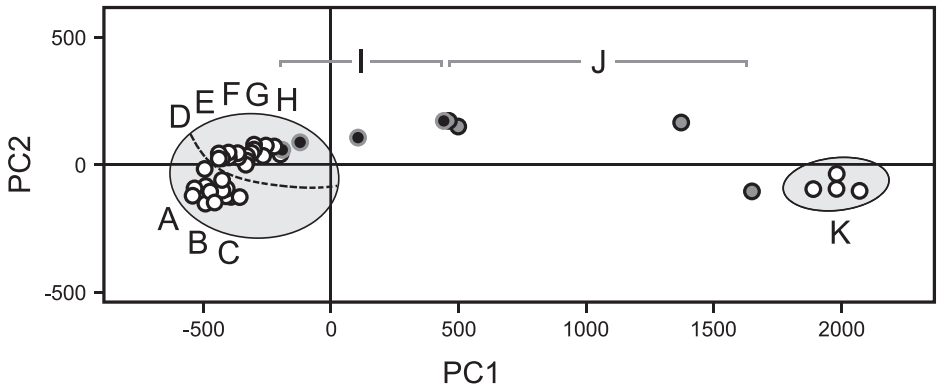


Figure 2. Principal component analysis (PCA) of overall gene expression variance. All 44 samples (4 per time point) are depicted in this figure. PC1 (67.9% of total variance) separates time points with logarithmic growth (A to H) from time points with stationary growth (time point K; $p < 0.000054$). PC2 (8.0% of total variance) separates time points A, B and C from D to H as indicated by the dashed line (not significant).

Functional categorization of significantly regulated genes

The genes from expression groups I, II and III in figure 3A were mapped to the biological pathway classification provided for *Bordetella pertussis* by the KEGG database[19,20]. The genes of the *virulence core regulon*[4] were included as an additional pathway. Expression group IV was disregarded, due to the inconsistent expression patterns of the 8 genes in this group. Mapping of 697 remaining genes in the other expression groups revealed that 25 pathways were significantly more abundant in at least one of the three groups. Compared to random distribution, these 25 pathways contained a significantly higher number of the regulated genes, 17 of which are depicted in figure 3B. The other 8 pathways had high overlap in gene content or function with the pathways in this figure and are therefore not shown. (The supplementary data file (table B) contains a list of all 142 KEGG pathways and the p-values associated with increased abundancy.) In addition to several pathways involved in translation and energy metabolism, genes of the *virulence core regulon* were significantly more abundant in expression group II ($p < 0.00001$). After time point H, 33 of the 56 *virulence core regulon* genes (59%) became significantly down regulated, including *bvgS*, *bvgA*, type 2 fimbriae (*fim2*), pertactin (*prn*) and filamentous haemagglutinin (*fhaB*) genes. (The supplementary data (table C) file provides a list that shows which *virulence core regulon* genes were down regulated in this study.)

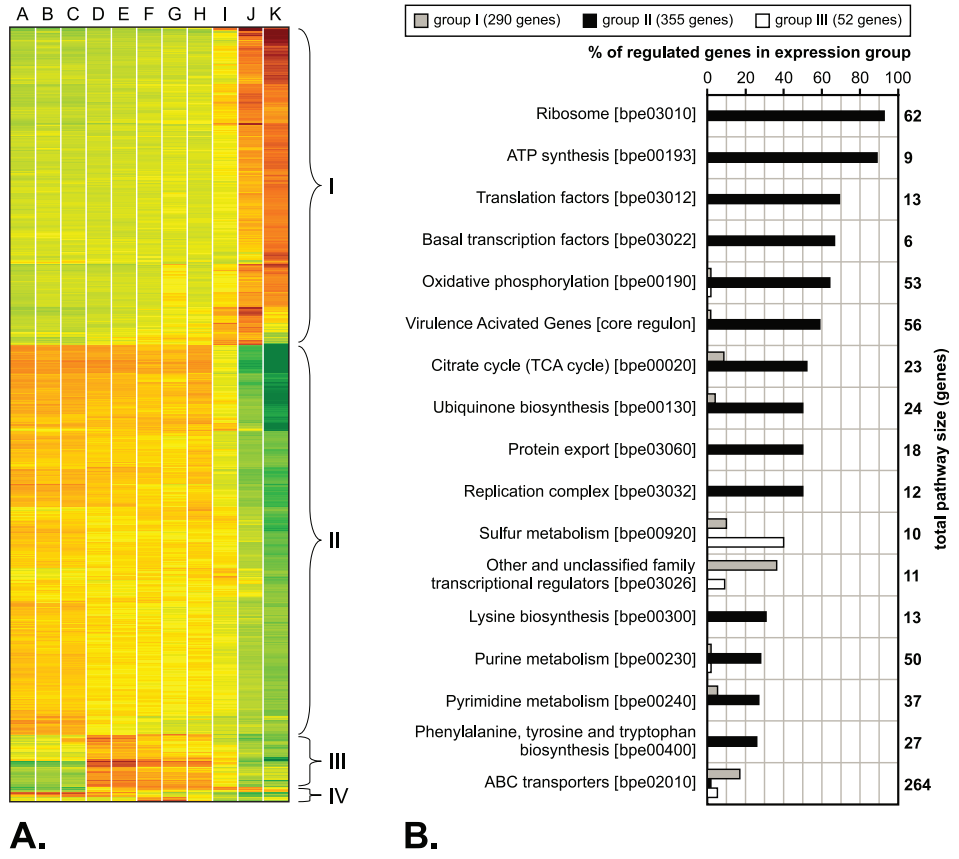


Figure 3. A: Clustering of 705 significantly regulated genes. Each row represents the relative expression of a single gene throughout the cultivation. Column names (A to K) correspond to the microarray time points in Figure 1. Genes are classified into four expression groups (I to IV). The color scale runs from strong down regulation (dark green), through no regulation (yellow), to strong up regulation (dark red).
B: Analysis of predominant KEGG pathways in expression groups I, II and III. Bold numbers on the right show the total number of genes in each pathway. Bars indicate which percentage of the total pathway size is found within each expression group. Grey bars correspond to expression group I, black bars to group II and white bars to group III. KEGG pathway ID's are given between brackets.

Product quality scores

The product quality scores were calculated for each time point, using the weighted gene expression values of the *virulence core regulon* genes in a PCA-derived model. This linear model includes all 56 genes of the *virulence core regulon*, because all these genes were previously found to be conserved between three *Bordetella pertussis* strains[4]. Furthermore, this model was developed to be applicable in future process development, when other subsets of the core regulon might be involved. Figure 4 shows the calculated product quality scores

on an arbitrary scale from 0.0 (lowest quality) to 10.0 (highest quality). These product quality extremes were derived from a bvg-modulation experiment, where a 10.0 score represents the average product quality of cultivations under normal, bvg(+) conditions and a 0.0 score is the average product quality obtained under modulated, bvg(-) growth conditions. The majority of product quality scores in this study was centered around an average of 10.0 (SD = 0.3) and includes all quadruplicate samples from time points A to I. Time point J has a lower average quality score of 8.9, as compared to time point A to I ($p < 0.075$; not significant) and time point K has a significantly lower product quality score of 6.2 ($p < 0.000094$), compared to time points A to I. (Table D of the supplementary data file contains all data that was used to construct the model for calculation of the product quality scores.) It shows the relationship between gene expression data, PCA scores, weight factor of each virulence core regulon gene and resulting product quality score.

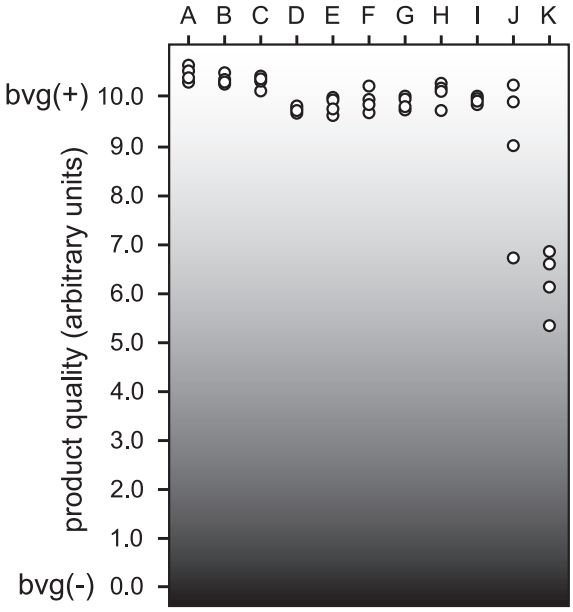


Figure 4. Product quality scores of microarray samples at different time points throughout the cultivation process. The gene expression of *virulence core regulon* genes was used to calculate product quality on an arbitrary scale (y-axis), where 0.0 corresponds to non-virulent conditions (bvg^-) that indicate a non-protective vaccine and 10.0 corresponds to virulent (bvg^+) conditions, that indicate optimal vaccine potency. White circles represent all samples taken during the four cultivations, the corresponding time points are globally indicated on the x-axis. The samples show a clear drift from virulent to less virulent conditions, which indicates that product quality is at risk towards the end of the cultivations (time point K).

Discussion

The bioreactor cultivation method used for this work, with chemically defined growth medium[10,11] and other previously described improvements[4,21], resulted in a highly reproducible process in terms of gene expression, bacterial growth and consumption of nutrients (Figure 1). Principal Component Analysis (PCA)[16,17] confirmed overall gene expression pattern similarity of samples at identical time points and revealed a significant shift of this pattern towards the end of the cultivation. The first principal component (PC1) in figure 2 separate samples taken during logarithmic growth (time points A to H) from samples from the stationary phase (time point K; $p < 0.0066$), through the transitional samples at time points I and J. Having established robust starting conditions for the cultivation of *Bordetella pertussis* in previous work[4], this PCA trajectory now provides a reliable process fingerprint for the selection of an optimal harvest point that balances between good quality and maximal yield.

After inoculation, cells entered logarithmic growth that became stationary upon limitation of the primary carbon and nitrogen sources (lactate and glutamate, respectively). This nutrient limitation induced a gene expression shift by up regulating biosynthesis-related ABC transporter genes (Figure 3, expression group I) and down regulating many genes from translation and energy metabolism pathways (Figure 3, expression group II). Interestingly, lactate and glutamate limitation also induced significant down regulation of 33 *virulence core regulon* genes[22,4], 28 of which were previously found to be growth phase and nutrient limitation associated by Nakamura *et al*.[23]. (The supplementary data file (table E) shows a Venn diagram of the overlap between the genes found in this study and by Nakamura *et. al*.) This observation has implications for the quality of the bacterial suspension in its use for a whole cell vaccine.

The use of full genome DNA microarray technology allows investigation of how an organism reacts to the continuously changing environment during batch fermentation processes like the one presented in this study[24,25]. Even though nutrient and biomass concentrations changed substantially between time points A and H, the overall gene expression pattern remained largely unchanged. The genes in expression group III are an exception, but this group is relatively small, and the effect was transient, producing a non-significant in PCA shift and did not involve *virulence core regulon* genes. The gene expression profile in all groups did change significantly after time point H, when nutrients became limiting. The resulting list of down regulated *virulence core regulon* genes contains many transcripts encoding virulence proteins that are known to be important for a protective immune response against *Bordetella pertussis*, like filamentous haemagglutinin (FHA), pertactin (Prn) and type 2 fimbriae[26,27] (Fim2). (See the supplementary data file (table C) for the full list of all core regulon genes.)

Differential mRNA expression in prokaryotes is more predictive for changes in protein expression than in eukaryotes, because eukaryotes modify their proteins after translation to a much higher extent. For this reason, the observed down regulation of 33 *virulence core regulon* genes is likely to be closely reflected in the protein composition of the whole cell pertussis vaccine that is made from these bacteria and is therefore useful as an predictor for vaccine quality. This hypothesis is supported by other studies, that reported a correlation between vaccine potency and virulence protein content[28] and a lower protective potency of vaccines that were harvested during the stationary growth phase[8].

This proposed link between bacterial virulence and vaccine potency was used to develop a statistical PCA-derived model for product quality measurement, which condenses the gene expression of all known virulence genes into a single number, the product quality score (Figure 4). This score allows benchmarking of a cultivation sample to a linear 0 to 10 scale for expected vaccine potency. The presented results show a clear correlation between the product quality score and nutrient availability, resulting in a significantly lower product quality score towards time point K ($p < 0.000094$). For process optimization studies, such a score is crucial when the effects of small changes in process conditions need to be distinguished and assessed for impact on product quality. Especially when the mandatory potency test is as crude and unreliable as in this case, the need for a quantitative quality scale is obvious, not only for process development, but also for clinical testing, where detailed information on the potency of clinical batches can be used to interpret the clinical responses more accurately[29].

Conclusions

To assure maximal quality and potency of the whole-cell vaccine, *Bordetella pertussis* batch fermentations should be harvested during logarithmic growth, before the critical concentration of 2 mM of either lactate or glutamate is reached, which corresponds to time point H. In this way rational, biologically relevant criteria are used to determine the optimal harvest point in respect to both yield and product quality. After determining the starting conditions in previous work[4], this study now provides a rational, science based harvest point for the batch fermentation process of *Bordetella pertussis* in whole-cell vaccine production.

DNA microarrays can be a reliable tool for assessing process performance and generating process understanding, but are labor intensive and can only be used to test a limited number of samples. Implementation of typical Process Analytical Technology (PAT)[9] sensors such as near infrared spectroscopy (NIRS) during batch fermentation, would generate a continuous, multivariate process fingerprint of the inoculation-to-harvest trajectory. On-line monitoring of nutrient concentrations for automated detection of the harvest point is also possible [30,31,32]. Such a science-based PAT approach, in combination with the presented model to measure product quality, could result in a process that yields whole-cell vaccines of such intrinsic quality and consistency[33] that it would change the perspective for the use of the Kendrick potency test[34,35]. Not only is this test cruel for the animals, it is also highly variable and gives yes or no answers at best (i.e. protection or no protection)[36]. With growing scientific confidence in the more consistent quality of vaccines produced with PAT controlled processes, the need for animal release tests on every single batch will eventually decline[9]. Additional research that validates the correlation between gene expression-based product quality scores and the potency of vaccines as measured with mandatory animal tests, could provide the scientific basis needed to take the next step in this direction.

Acknowledgements

The authors gratefully acknowledge Dr. B. Zomer for the NMR nutrient analyses. This work was partly funded by Senter Novem, The Netherlands, grant number TSGE3067.

References

1. Blencke HM, Homuth G, Ludwig H, Mader U, Hecker M, Stulke J. 2003. Transcriptional profiling of gene expression in response to glucose in Bacillus subtilis: regulation of the central metabolic pathways. Metab Eng 5(2):133-49.
2. Nystrom T. 2004. Stationary-phase physiology. Annu Rev Microbiol 58:161-81.
3. Yoon SH, Han MJ, Lee SY, Jeong KJ, Yoo JS. 2003. Combined transcriptome and proteome analysis of Escherichia coli during high cell density culture. Biotechnol Bioeng 81(7):753-67.
4. Streefland M, van de Waterbeemd B, Happe H, van der Pol LA, Beuvery EC, Tramper J, Martens DE. 2007. PAT for vaccines: The first stage of PAT implementation for development of a well-defined whole-cell vaccine against whooping cough disease. Vaccine.
5. Nakamura MM, Liew SY, Cummings CA, Brinig MM, Dieterich C, Relman DA. 2006b. Growth phase- and nutrient limitation-associated transcript abundance regulation in Bordetella pertussis. Infect.Immun. 74(10):5537-5548.
6. Cummings CA, Bootsma HJ, Relman DA, Miller JF. 2006a. Species- and strain-specific control of a complex, flexible regulon by Bordetella BvgAS. J.Bacteriol. 188(5):1775-1785.
7. Mills KH. 2001. Immunity to Bordetella pertussis. Microbes.Infect. 3(8):655-677.
8. Bellalou J, Relyveld EH. 1984a. Studies on culture conditions of Bordetella pertussis and relationship to immunogenicity of vaccines. Ann.Microbiol.(Paris) 135B(1):101-110.
9. FDA. 2004. Guidance for Industry, PAT - a framework for innovative pharmaceutical development, manufacturing and quality assurance. Pharmaceutical CGMP's.
10. Thalen M, van den IJssel J, Jiskoot W, Zomer B, Roholl P, de Gooijer C, Beuvery C, Tramper J. 1999. Rational medium design for Bordetella pertussis: basic metabolism. J Biotechnol 75(2-3):147-59.
11. Thalen M, Venema M, van den IJssel J, Berwald L, Beuvery C, Martens D, Tramper J. 2006. Effect of relevant culture parameters on Pertussis Toxin expression by Bordetella pertussis. Biologicals 34(3):213-20.
12. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MT, Churcher CM, Bentley SD, Mungall KL and others. 2003. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. Nat Genet 35(1):32-40.
13. Kasahara T, Miyazaki T, Nitta H, Ono A, Miyagishima T, Nagao T Urushidani T. 2006. Evaluation of methods for duration of preservation of RNA quality in rat liver used for transcriptome analysis. J Toxicol Sci 31(5):509-19.
14. Mutter GL, Zahrieh D, Liu C, Neuberg D, Finkelstein D, Baker HE, Warrington JA. 2004. Comparison of frozen and RNALater solid tissue storage methods for use in RNA expression microarrays. BMC Genomics 5(1):88.
15. Ramalho AS, Beck S, Farinha CM, Clarke LA, Heda GD, Steiner B, Sanz J, Gallati S, Amaral MD, Harris A and others. 2004. Methods for RNA extraction, cDNA preparation and analysis of CFTR transcripts. J Cyst Fibros 3 Suppl 2:11-5.
16. Quackenbush J. 2001. Computational analysis of microarray data. Nat Rev Genet 2(6):418-27.
17. Raychaudhuri S, Stuart JM, Altman RB. 2000. Principal components analysis to summarize microarray experiments: application to sporulation time series. Pac Symp Biocomput:455-66.
18. Kredich NM. 1971. Regulation of L-cysteine biosynthesis in Salmonella typhimurium. I. Effects of growth of varying sulfur sources and O-acetyl-L-serine on gene expression. J Biol Chem 246(11):3474-84.
19. Nakao M, Bono H, Kawashima S, Kamiya T, Sato K, Goto S, Kanehisa M. 1999. Genome-scale Gene Expression Analysis and Pathway Reconstruction in KEGG. Genome Inform Ser Workshop Genome Inform 10:94-103.

20. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 27(1):29-34.
21. Thalen M, van der Ark A, van den IJssel J, van Straaten I, Jansen D, Beuvery C, Martens D, Tramper J. 2008. Improving the cellular pertussis vaccine: Increased potency and consistency. Vaccine 26(5):653-63.
22. Cummings CA, Bootsma HJ, Relman DA, Miller JF. 2006b. Species- and strain-specific control of a complex, flexible regulon by Bordetella BvgAS. J Bacteriol 188(5):1775-85.
23. Nakamura MM, Liew SY, Cummings CA, Brinig MM, Dieterich C, Relman DA. 2006a. Growth phase- and nutrient limitation-associated transcript abundance regulation in Bordetella pertussis. Infect Immun 74(10):5537-48.
24. Elias DA, Tollaksen SL, Kennedy DW, Mottaz HM, Giometti CS, McLean JS, Hill EA, Pinchuk GE, Lipton MS, Fredrickson JK and others. 2007. The influence of cultivation methods on Shewanella oneidensis physiology and proteome expression. Arch Microbiol.
25. Gupta P, Lee KH. 2007. Genomics and proteomics in process development: opportunities and challenges. Trends Biotechnol 25(7):324-30.
26. Corbel MJ, Xing DK. 2004. Toxicity and potency evaluation of pertussis vaccines. Expert Rev Vaccines 3(1):89-101.
27. Kinnear SM, Marques RR, Carbonetti NH. 2001. Differential regulation of Bvg-activated virulence factors plays a role in Bordetella pertussis pathogenicity. Infect Immun 69(4):1983-93.
28. Hamstra HJ, Kuipers B, Schijf-Evers D, Loggen HG, Poolman JT. 1995. The purification and protective capacity of Bordetella pertussis outer membrane proteins. Vaccine 13(8):747-52.
29. Cogdill R, Drennen J. 2008. Risk-based Quality by Design (QbD): A Taguchi Perspective on the Assessment of Product Quality, and the Quantitative Linkage of Drug Product Parameters and Clinical Performance. Journal of Pharmaceutical Innovation 3(1):23-29.
30. Arnold SA, Gaensakoo R, Harvey LM, McNeil B. 2002. Use of at-line and in-situ near-infrared spectroscopy to monitor biomass in an industrial fed-batch Escherichia coli process. Biotechnol Bioeng 80(4):405-13.
31. Ferreira AP, Alves TP, Menezes JC. 2005. Monitoring complex media fermentations with near-infrared spectroscopy: comparison of different variable selection methods. Biotechnol Bioeng 91(4):474-81.
32. van Sprang ENM, Streefland M, Ramaker H-J, van der Pol LA, Beuvery EC, Smilde AK. 2007. Manufacturing Vaccines: An Illustration of Using PAT Tools for Controlling the Cultivation of Bordetella pertussis. Quality Engineering 19(4):373 - 384.
33. Gnoth S, Jenzsch M, Simutis R, Lubbert A. 2007. Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control. J Biotechnol.
34. Irwin JO, Standfast AF. 1957. The intracerebral mouse-protection test for pertussis vaccines. J Hyg (Lond) 55(1):50-80.
35. Kendrick PL, Eldering G, Dixon MK, Misner J. 1947. Mouse Protection Tests in the Study of Pertussis Vaccine: A Comparative Series Using the Intracerebral Route for Challenge. Am J Public Health Nations Health 37(7):803-10.
36. Xing D, Das RG, O'Neill T, Corbel M, Dellepiane N, Milstien J. 2001. Laboratory testing of whole cell pertussis vaccine: a WHO proficiency study using the Kendrick test. Vaccine 20(3-4):342-51.

CHAPTER 5

Manufacturing vaccines: An Illustration of Using PAT Tools for Controlling the Cultivation of *Bordetella pertussis*

Eric N.M. van Sprang[a], Mathieu Streefland[b], Leo A. van der Pol[b], E. Coen Beuvery[c], Henk-Jan Ramaker[a] and Age K. Smilde[d,e],

[a]TIPb, Applied Industrial Process Control, Kruislaan 419, 1098 VA, Amsterdam, The Netherlands.
[b]Netherlands Vaccine Institute (NVI), Unit Research and Development, PO Box 457, 3720 AL Bilthoven, The Netherlands
[c]PAT-Consultancy, Kerkstraat 66, 4132 BG Vianen, The Netherlands
[d]University of Amsterdam, Biosystems Data Analysis group, Nieuwe Achtergracht 166, 1018 WV, Amsterdam, The Netherlands
[e]TNO Quality of Life, Utrechtseweg 48, 3700 AJ Zeist, The Netherlands

Quality Engineering (2007) 19(4): 373-384

Abstract

An illustration of the operational consistency of the upstream part of a biopharmaceutical process is given. For this purpose four batch cultivations of *Bordetella pertussis* have been executed under identical conditions. The batches have been monitored by means of two fundamentally different process sensors. First, common single channel measurements such as temperature, pH, dissolved oxygen (DO) and flow rates are used and secondly the multi-channel measurements from the NIR (Near Infrared) analyzer. Because of the fundamental differences between the two types of measurements, two models have been developed to evaluate the operational consistency. The last sensor studied is a typical representative of process analyzers which are described in the PAT (Process Analytical Technology) guidance document issued in 2004 by the American Food and Drug Administration (FDA). Data from both sensors have been evaluated by a multivariate tool for data acquisition. This resulted in two different performance models. Again this approach is characteristic for the implementation of PAT for the manufacture of biopharmaceuticals.

With both performance models, we were able to explore the operational consistency of the batches. In addition, the performance models were also able to detect a deviating batch. Furthermore, it was shown that both sensor types gave partly overlapping information since a deviation in the batch profiles of the logged process variables was accompanied by a deviation in the spectral batch profiles.

The performance models are valuable tools in developing advanced monitoring and control systems for biopharmaceutical processes. Using such models, advanced knowledge based systems can be developed to detect abnormal situations in an early stage and remove the cause.

The way of data processing described in this paper is relatively new in the biopharmaceutical industry. The NIR analyzer and both performance models presented in this paper are clear ingredients for better process understanding and process control, as intended in the FDA's PAT Initiative. This initiative is part of the FDA's strategy of cGMP (current good manufacturing practice) for the 21st Century and aims at introducing innovations in both the manufacturing of biopharmaceuticals and the development of new biopharmaceuticals.

This study shows the feasibility of two typical PAT tools for controlling the manufacturing of biopharmaceuticals. To the best of our knowledge such feasibility study is not documented up to now in the scientific literature.

Introduction

The production of most biopharmaceuticals such as *Bordetella pertussis* vaccine is usually performed in a batch wise manner. Batch processes are characterised by a recipe driven process. That is, they have a clear starting point, an end-point and non-linear trajectories.

Biochemical processes as the cultivation of *Bordetella pertussis,* are driven by the underlying biochemical reactions. As a result there exists a dynamic system with a predictable outcome. However, cultivation processes are liable to all kinds of variations that can influence the large number of simultaneous reactions and processes[1]. These variations can strongly influence the processes in a negative manner, leading to unpredictable behaviour. The approach of controlling and monitoring (bio)-pharmaceutical processes to a desired end-point is encouraged by regulatory authorities and described in the FDA's Guidance for Industry on Process Analytical Technology (PAT). A full version of the document can be downloaded from the FDA's website *www.fda.gov/cder/gmp/gmp2004/guidance/6419fnl.pdf.*

The goal of the FDA's PAT initiative is to break the barrier that strict regulation has formed against flexible processing and to bring the pharmaceutical industry to the current state of the art of industrial processing and as already indicated the introduction of innovation in the manufacturing and development of biopharmaceuticals.

While batch production runs are performed according to a recipe, experience learns there can be differences between identically operated batches. These differences become apparent in the profiles of the logged process variables and in the end product quality. The cause of these differences can often be found in variations in the starting materials used for the cultivation, operating conditions or manual interventions by the operators during the execution of the cultivation step. For the production of biopharmaceuticals such as vaccines, it is therefore very important to monitor and control process variations and prevent and reduce unknown or abnormal process situations in order to reduce the non-acceptable risks inherent to this more 'classical' way of manufacturing biopharmaceuticals.

To monitor the evolution of a batch process, it is common to take samples during the operation of a batch for (bio)-chemical analysis. Furthermore, modern process hardware is equipped with single channel sensors such as pressure or temperature sensors. In addition, industrial processes are equipped with multi-channel sensors such as spectroscopic process analysers. Both sensors are very informative for the current state of the process but they are fundamentally different in nature. This results in different, yet complementary types of process information.

The basic idea behind consistency analysis is that a guaranteed product quality is obtained by producing in a consistent manner. It is a priori not always known what the effects of variations are on the end-product quality. It is therefore important to work proactively and monitor and secure the critical attributes of the production chain. However, before developing monitoring and control applications, it is important to understand the variations present in the data.

To understand the variation present in a production process historical process data can be collected and analysed. In practice, it is often the case that many operational process changes have taken place such that the historical data is not representative for the current manner of production. However, although changes have been made, the historical data will give insight on the working of the process and will help to identify bottlenecks. Based on such an analysis critical attributes can be identified and used for controlling and securing the process.

The ultimate goal is a well controlled process with a guaranteed end-product quality. This can only be realized by monitoring the critical process attributes for the entire production chain. It can be expected that if all critical process attributes operate within predefined or statistical control limits, a guaranteed good quality product is produced. Besides the determination of the critical process attributes and derivation of acceptable limits of operation a first start is to study the current way of operating each unit operation on a pilot scale.

The aim of this paper is to illustrate a method for processing of data from both single and multi-channel sensors for assessment of process consistency on a pilot scale. The described process and process measurements are the first step of an optimization program. The experiments are expensive, therefore, prior to performing experiments according to an experimental design it should be verified wheter the process can be operated stable around a process mean.

To verify the operational consistency of the cultivation, four batch cultivation runs of *Bordetella pertussis* bacteria have been operated in an identical way. The cultivations are carried out in a bioreactor that is equipped with standard process sensors (temperature, DO and pH) and a near infrared (NIR) sensor.

The bacterial cultivation of the Bordetella pertussis

Bordetella pertussis bacteria were cultured on the chemically defined THIJS medium Thalen *et al*.[2], which is designed to meet the bacterium's nutritional demands. Using such a chemically defined medium eliminates variation that is commonly caused by undefined medium components such as yeast extract. The medium contains two essential nutrients,

lactate and glutamate. Glutamate is essential for growth while lactate enhances the growth rate.

All bioreactor experiments were carried out in a 7 L *in situ* sterilizable glass bioreactor at a working volume of 4 L. Temperature, dissolved oxygen (DO) concentration, pH and stirrer speed were controlled at 34°C, 30%, 7.2 and 450-650 rpm, respectively. Oxygen concentration was first controlled with increments of 50 rpm in stirrer speed until 650 rpm was reached. Subsequently, DO was controlled by increasing the fraction of oxygen in the headspace. After the bioreactors were filled with 4L THIJS medium and had reached setpoint temperature and DO=100%, they were inoculated with 200 ml preculture at optical density at 590 nm (OD_{590})=1.0±0.05. Samples were taken for OD_{590} and at line nutrient concentration measurements (lactate and glutamate) using a YSI 2700 analyzer (Yellow Springs Instruments, Yellow Springs, USA). Samples were sterile filtered (0.22 μm) and stored at -20°C for later NMR (Nuclear Magnetic Resonance) analysis for confirmation of the nutrient concentrations. The bacteria were cultured until a final optical density of OD_{590}=1.5±0.1 was reached.

Process information from the two different types of sensors

Prior to discussing the mathematical tools for the explorative analysis, it is important to realize that in the current analyses there are two distinct types of process data. The first type is the commonly known as 'engineering variables' or single channel sensors. The engineering variables usually consist of physical process information such as dissolved oxygen concentration, temperature and pH. The second type of process data is multi-channel data from process analyzers such as near-infrared (NIR) spectroscopy. Data obtained from a NIR process analyzer is rich in biochemical information and characterised by a high degree of collinearity between the variables in the data.

Within the single channel measurements, there is a distinction in nature of logged process variables; first, there are discrete variables such as set-points and valve settings, and second, there are the dynamic (continuous) variables such as temperatures or viscosity.

Extraction of chemical information from NIR spectra requires experience with such a process analyzer. NIR spectrometers acquire spectra in the range between 12000 cm^{-1} and 4000 cm^{-1}. NIR-active components such as the substrates, which absorb in this region, are characterized by broad overlapping peaks. This implies that there is no clear and unique region for one particular component in the medium during the cultivation of the bacterium, but the whole spectral fingerprint is important. To be able to relate this complex medium to concentrations of the substrates, samples have been analyzed by the analytical reference methods such as

the YSI, Nuclear Magnetic Resonance (NMR) and Optical Density (OD_{590}). Next, models are developed to relate the spectra to the results of the analytical reference methods.

Consistency analysis

Consistency analysis based on single channel measurements

Using logged process variables, much insight about the process is gained by simply presenting these process variables as time series. In this relative simple manner features such as process dynamics and process disturbances can be identified. To give an example, the time series of the logged process variables are presented in Figure 1. For the three controllers active in the bioreactor control system (pH, DO and Temperature) all controller outputs (CO) and control actions (acid flow, air flow, etc) are plotted besides the actual measurements of the parameters. This does not only provide insight on the performance of a certain controller, but also on the output which the controller had to deliver to keep its setpoint. This is valuable information specific for each process run.

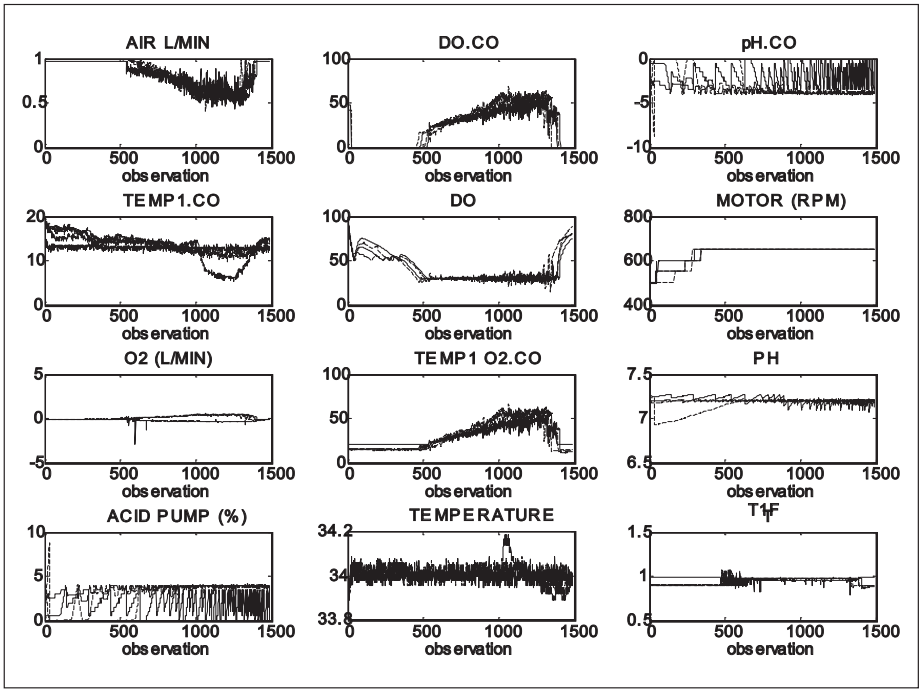


Figure 1. Time series for the logged process variables of the four consistency runs.

However, if the number of batches and logged process variables increases, which is often the case in an industrial environment; it becomes difficult to study these high dimensional data. A tool often used in industry to study a high dimensional dataset is a

subspace model constructing using principal component analysis, PCA[3]. The aim of a subspace model is to capture the systematic information of a high dimensional dataset consisting of collinear data (such as industrial process data) and project it onto a lower dimensional space without loss of information. Next, the lower dimensional space is explored.

There are many ways to model an industrial batch process using subspace models. The choice of the model depends for a large extent on the aim of the experimenter and the availability of historical data. In the early nineties the first strategies to study the batch consistency were proposed[4,5,6]. Following the ideas of Nomikos and MacGregor several alternatives were suggested to model the data. A detailed overview of possible batch modelling strategies can be found in Lee and Vanrolleghem[7] and Van Sprang *et al.*[8]. In addition, a more detailed discussion and applications of these methods in a PAT context can be found in Kourti[9], Lopes *et al.*[10] and Wold *et al.*[11].

Consistency analysis of single channel measurements

As stated earlier, the focus of this study is to determine if the cultivation step is reproducible in terms of operational conditions prior to executing designed experimental batch runs. As a result there are only four batch runs available. Since there are only four runs available it is impossible to make any statements based on statistical arguments. However, by comparing the individual trends in a reduced space a fair impression of the operational consistency is obtained.

A series of batch runs can be presented as a three-way array $\underline{\mathbf{X}}$ (I x J x K), where I is the number of batches, J the number of process variables and K the number of observations. The two often used approaches to model the three-way array are the approaches suggested by Nomikos and MacGregor[4] and Wold *et al.*[6].

In the approach of Nomikos and MacGregor, the three-way array $\underline{\mathbf{X}}$ is unfolded in the batch direction leading to the matrix $\mathbf{X}$ (I x JK). Next, $\mathbf{X}$ (I x JK) is modelled using a PCA model with R latent variables as shown in equation (1).

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \qquad (1)$$

where the loadings $\mathbf{P}$ (JK x R) capture all systematic behavior. The score matrix $\mathbf{T}$ (I x R) represents the differences and similarities between the batches with respect to this systematic behavior and matrix $\mathbf{E}$ captures the unmodeled variation. The model of eq (1) is fitted using a least-squares approach, hence, by minimizing the sum-of-squared residuals of $\mathbf{E}$.

In the approach of Wold *et al.*[6], the three-way array $\underline{\mathbf{X}}$ is unfolded in the variable direction leading to the matrix $\mathbf{X}_w$ (*KI* x *J*). Next, $\mathbf{X}_w$ (*KI* x *J*) is modelled using a PCA model with *R* latent variables as shown in equation (2).

$$\mathbf{X}_w = \mathbf{T}_w\mathbf{P}^T_w + \mathbf{E}_w, \quad (2)$$

where the loadings $\mathbf{P}_w$ (*J* x *R*) capture the overall systematic behavior between the variables. The score matrix $\mathbf{T}_w$ (*KI* x *R*) captures the dynamic behaviour of the batches in the model plane and matrix $\mathbf{E}_w$ captures the unmodeled variation.

Note that the proposed strategies assume that all batches have time trajectories of equal run length. If that is not the case, synchronization strategies such as dynamic time warping must be used to align the batches[12]. The essential difference between (1) and (2) is the way the batches are treated. In (1) the behaviour of a batch is summarized in *R* numbers (one row of T), whereas in (2) the behaviour is summarized in *K* x *R* numbers (the part of T_w related to the specific batch). Loosely stated, in (1) each batch is considered as one observation, whereas in (2) each time-point in one batch is considered as one observation.

Process data is usually measured in different units. The difference in units can give certain process variables an unfair weight of importance in the models. Therefore, prior to modelling process data should be standardized (also referred to as pre-processing). The choice of pre-processing is depending on the aim of the analysis. Sometimes it is desired to give certain process variable more or less weight depending on the interest of the analysis. It is important to realize that the results of the analysis can be affected by the choice of pre-processing. When working with process variables it is common to standardize the data to unit standard deviation and mean zero, which is referred to as autoscaling.

If a relative high number (~ 40) of batches is available for analysis the method as proposed by Nomikos and MacGregor[4], is well suited to study the batch consistency. Similar batches will be clustered together in the model plane and the loadings can be explored for correlations between process variables. As stated earlier, the current dataset consists of four batches which are assumed to be identical. Using the approach of Nomikos and MacGregor[4] with only four observations is not very informative since this leads to four points in an *R* dimensional space. Furthermore, there is not enough data to estimate the model parameters properly. Hence the approach of Nomikos and MacGregor[4] is not used in the current analysis.

A possible model strategy that can be used for the current dataset is a leave one batch out version of the method suggested by Wold[6]. The principle idea is to leave one batch out of the set of available batches and use the three remaining batches to develop a model. Next,

the left out batch is projected onto the model and the obtained score trajectories and model residuals are used for exploring the operational batch behaviour. This process is repeated until all batches are left out once and projected on the model built with the remaining ones. Since the batches are operated under identical conditions it is expected that similar profiles are found for each batch run.

In the present study, there are four experimental batches operated under identical conditions including the termination of a batch. The batches are denoted as batch 11, batch 12, batch 13 and batch 15. Each batch has a total of $K = 1490$ observations and $J = 12$ logged process variables as presented in Figure 1. The data are modelled using the leave one batch out approach. Prior to modeling, the data is column mean centered and divided by its standard deviation. The number of latent variables is determined using a full-cross validation. In total four models are developed. The explained variation per model is presented in Table 1.

Table 1. Percentage explained variation per model using a leave one out strategy for the modelling approach suggested in Wold *et al*.[18]

Batch not included	PC 1 percentage explained variation	PC 2 percentage explained variation	PC 3 percentage explained variation	Total percentage explained variation
Batch 15	50.4	15.4	9.4	75.2
Batch 11	49.0	16.1	9.2	74.3
Batch 12	61.4	14.9	7.8	84.1
Batch 13	47.5	16.3	9.9	73.7

Table 1 shows that leaving out batch 12 increases the amount of explained variation with an equal number of latent variables compared with the other models. This is an indication that the batch 12 has a different behaviour.

The scores $\mathbf{T}_{lo}$ of the left out batches are found by projecting the observations onto the model plane $\mathbf{P}_{lo}$ according to:

$$\mathbf{t}^{T}_{k,lo} = \mathbf{x}^{T}_{k,lo}\mathbf{P}_{lo} \tag{3}$$

By presenting the new scores $\mathbf{t}_{k,lo}$ obtained after projecting the observations for the left out batches the batch evolution is easily visualized as can be seen in Figure 2 (a), (b) and (c). There, the dynamic behaviour between the four batches is compared. The corresponding residuals are found by:

$$\mathbf{e}_{k,lo} = \mathbf{x}_{k,lo} - \mathbf{P}_{lo}\mathbf{t}_{k,lo} \tag{4}$$

Next, the trajectories of the sum of squared residuals (SSE) are visualized as time series in Figure 2 (d).

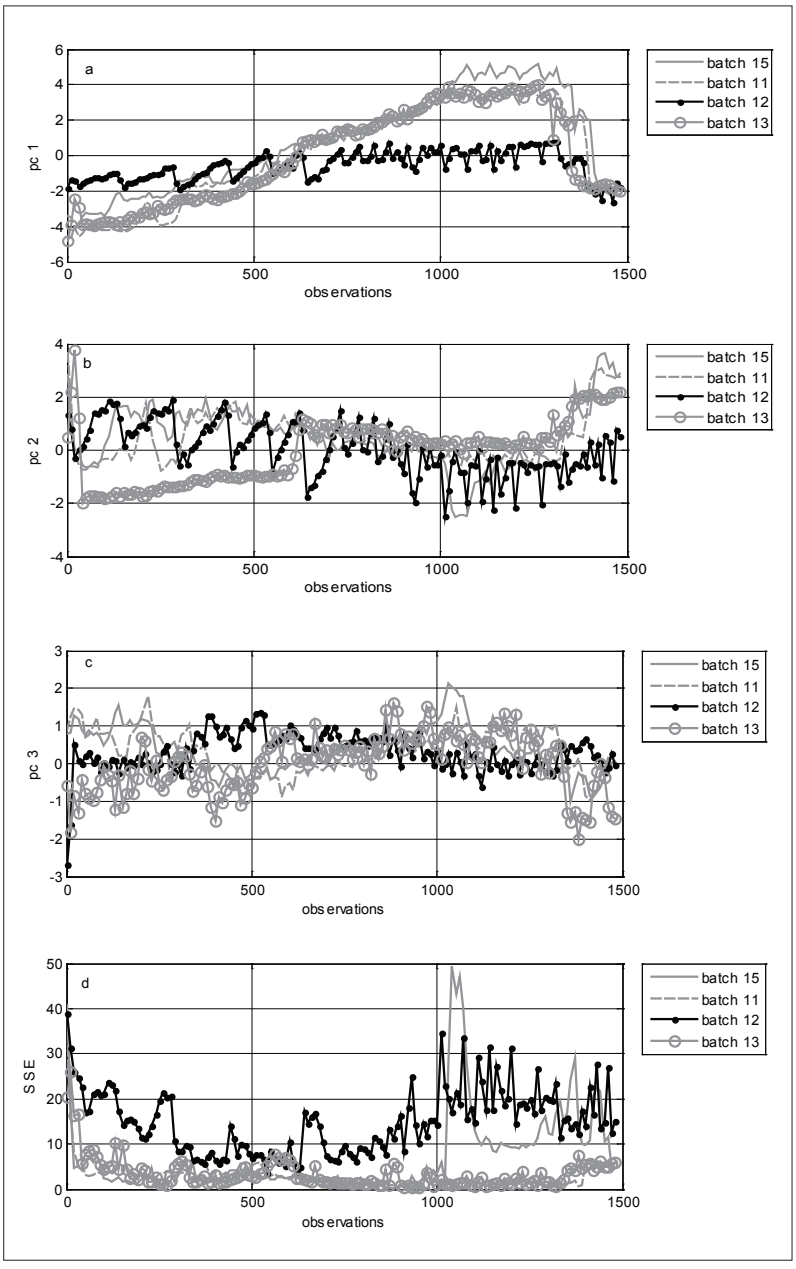


Figure 2. Results of monitoring the batch profiles using PCA. The dynamic behaviour for the first latent variable (a). The dynamic behaviour of the second latent variable (b). The dynamic behaviour of the third latent variable (c). The unmodeled variation (d).

Figure 2 shows the dynamic behaviour of the four batches presented in the first three latent variables. The unmodeled variation is captured in the SSE. The first latent variable in Figure 2 (a) shows that Batch 11, Batch 13 and Batch 15 follow a similar trend. However, Batch 12 follows a distinct pattern both in the dynamic behaviour in the first component and in distance from the model plane (SSE) Figure 2 (d). The second latent variable in Figure 2 (b) shows that the batches suffer from large variations but seem to follow a similar trend except for batch 13 that shows a deviating pattern in the first 500 observations during its evolution. Further analysis learns that this behaviour corresponds to a deviating pH signal for batch 13.

The third latent variable is not very informative. The unmodeled variation captured by the SSE shows a higher contribution over time for batch 12. The behaviour of batch 12 was unexpected; a more detailed analysis learned that the pH control during the operation of Batch 12 was problematic. This is due to the settings of the control system. Apparently, the required control action is not sufficient to keep the process under control or a stronger control action was required. Furthermore, a high contribution is observed for batch 15 starting from observation 1000. Further analysis learned that this behaviour is related to the temperature behaviour of this batch.

Consistency analysis using multi-channel data (NIR spectra)

To relate measured NIR spectra to the attributes of interest (lactate, glutamate and the optical density) a calibration model is constructed. Because of the nature of NIR spectra (the presence of collinearity and overlapping peaks) a multiple regression model is constructed usually based on Principal Component Regression (PCR) or Partial Least Squares (PLS).

During the development of a multivariate calibration model for use within an industrial environment a number of challenges are faced. The complexity of the sampling matrix and low concentration levels (~1 mmol/l) requires some effort. It is very important to have representative samples with the attributes of interest within the sample matrix. Although this may seem trivial, it often happens that due to a lack of scrutiny, samples are collected and presented for calibration that are not representative for samples under operational conditions. In addition, there should also be enough analyte signal in the spectra to create a calibration model. The latter is enhanced by spectral data processing strategies[13]. The choice and type of transformation strongly depends on the used sampling interface such as measurement type (reflection, transmission) and measurement environment (the sampling matrix).

In some situations, it might be the case that there are no reference measurements available to create a calibration model. Then, it is possible to develop models based on process spectra only[14] and use those directly for monitoring.

Assuming having a series of *I* batch processes that are monitored during their batch evolution using spectroscopy, for each observation *k* a spectrum with *J* channels is obtained. Next, the collected spectra can be arranged in a three-way array $\underline{\mathbf{X}}$ (*I* x *J* x *K*). Depending on the situation, there are in principle two choices.

The first choice is modeling the process spectra using a subspace model in a similar manner as discussed in the previous section. That is, the three-way array with collected process spectra $\underline{\mathbf{X}}$ (*I* x *J* x *K*) is unfolded in the variable direction resulting in a matrix $\mathbf{X}_s$ (*KI* x *J*), where *I* is the number of batches, *J* the number of channels and *K* the number of observations. Next, the data is modeled according to:

$$\mathbf{X}_s = \mathbf{C}\mathbf{S}^T + \mathbf{E}_s \tag{5}$$

where **C** (*KI* x R_s) captures the relative concentrations in a R_s dimensional space, and **S** (*J* x R_s) captures the spectral information of the absorbing components and defines the model plane in the R_s dimensional space.

The behaviour of new batches is studied by projecting observations $\mathbf{x}_{s,k}$ (*J* x 1) on the spectral model plane **S.** In this manner $\mathbf{c}_k$ is obtained. Next, the dynamic behavior of the batches can be studied by presenting the quantitative part as time series

The second choice is to incorporate a priori known process information into the model parameters. This can be done by putting restrictions on the model parameters during the development of the model. A more detailed discussion is given in Ramaker *et al.*[15]; van Sprang *et al*[16].

An example of the acquired spectra is given in Figure 3. The spectra are obtained with a Bruker Optics Matrix F NIR using 5 mm path length and a $1cm^{-1}$ resolution.

In Figure 3 the evolution is shown of a batch captured by the NIR analyzer. In this study there are three attributes of interest: 1) Lactate, 2) Glutamate and 3) the Optical density. For each of the attributes a multivariate calibration model is developed.

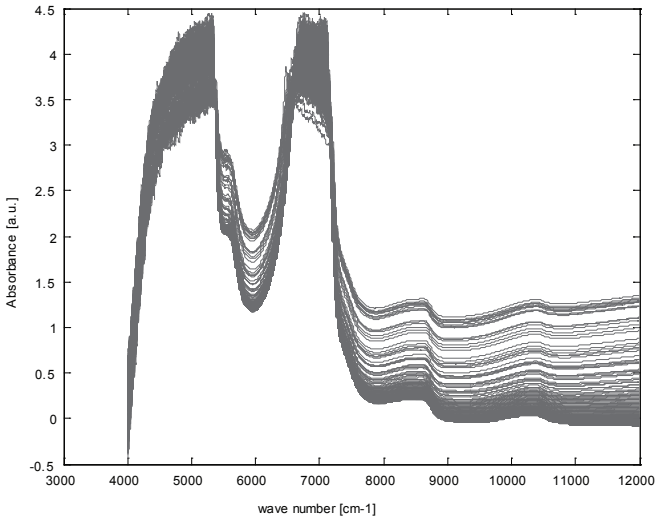


Figure 3. Raw spectra obtained for one batch run (batch 11). Spectra move from bottom to top during batch evolution.

Developing NIR calibration models using PLS

As stated previously, projection to latent structures (or partial least squares, PLS) is a commonly used regression technique in spectroscopy to regress a response variable y (one of the attributes of interest) on a set of predictors spectra **X** (*N* x *J*). PLS is a subspace regression method which maximizes the covariance between the scores **t,y**. This is done per dimension and finally collected in a matrix **T**. The matrix **X** is decomposed in three matrices **T, P** and **E**. The score matrix **T** (*N* x *R*) is made up of *R* score vectors. Matrix **P** (*J* x *R*) is loading matrix with *R* loading vectors and matrix **E** (*N* x *J*) consists of the model residuals. Vector **y** is decomposed in a weight matrix **q** (*R* x 1) and residual vector **f** (*N* x 1).The PLS decomposition is given in matrix notation in equation (6) to (8).

$$\mathbf{X}_s = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{6}$$

$$\mathbf{y} = \mathbf{T}\mathbf{q} + \mathbf{f} \tag{7}$$

$$\mathbf{T} = \mathbf{X}\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1} \tag{8}$$

where the matrix **W** (*J* x *R*) is an orthogonal weight matrix. Using the model parameters **W**, **P** and **q**, the response variable y is predicted on-line each time a new observation (*J* x 1) is measured according to equation 9 and equation 10.

$$\hat{\mathbf{t}}'_{new} = \mathbf{x}^{T}_{new}\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1} \tag{9}$$

$$\hat{\mathbf{y}} = \mathbf{t}^{T}_{new}\mathbf{q} \tag{10}$$

An approximate prediction interval for the individual response at (1- α)×100 confidence level is given by

$$\hat{\mathbf{y}} \pm \mathbf{t}_{v,\alpha/2}\hat{\sigma}(\hat{\mathbf{y}}\text{-}\mathbf{y}) \tag{11}$$

where $t_{v,\alpha/2}$ is the $(1\text{-}\alpha/2)^{th}$ quantile of a t-distribution with *v* degrees of freedom and $\hat{\sigma}(\hat{\mathbf{y}}\text{-}\mathbf{y})$ denotes the estimation of the standard error of prediction[17].

The calibration models were based on 52 samples, which are collected during the four batch runs. For each sample a spectrum is taken and analyzed accordingly by the analytical reference methods discussed in Section 2. The models are developed using an iterative approach. First, the spectra are screened to select informative spectral regions. This means that noisy regions from the data are removed. This first screening is primarily based on visual inspection of the spectra and theoretical knowledge about existing absorption bands. Next, the spectra are pre-processed by taking the second derivative spectra to remove unwanted variations from the spectra such as scatter effects or baseline shifts/drifts. The number of latent factors, or components, is determined by a full cross-validation. During this exercise outliers are removed from the data and a re-calibration is started. Then, the iterative procedure is stopped. Normally, model development can take several iterations in order to optimize the model. In the current study, the number of samples is not enough for a full model optimization and model validation but enough to generate models that give a reasonable prediction to monitor the dynamics of the attributes.

Operational consistency based on NIR spectra

The model settings and outcomes are presented in Table 2. From this table, it can be seen that the calibration error measured by the Root Mean Square Error of Calibration (RMSEC) for lactate and glutamate are approximately 1 mmol/l. Closer inspection of the model residuals reveal that the models have more difficulties to predict the lower lactate and glutamate (< 2 mmol/l) concentrations. Probably, the absorption signal of lactate and glutamate are too low at those concentrations. Normally, the calibration error of the cross validated models, expressed as the Root Mean Square Error of Cross Validation (RMSECV)[18], are approximately an order of 1.5 / 2 higher compared to RMSEC. The results show that this is indeed the case.

Table 2. Model settings and prediction results.

Explained variables	Latent variables	RMSEC (mmol/l)	RMSECV (mmol/l)	Selected spectral region (cm^{-1})
Lactate	7	1,1	1,6	5400 - 6300 & 7160 – 9000
Glutamate	3	0,9	0,8	5400 - 6300 & 7160 – 9000
Optical density	5	0.1	0.2	7160 – 9000

The batch profiles for lactate, glutamate and the optical density as predicted by the calibrated NIR sensor are presented in Figure 4.

Unfortunately, due to a software error it was not possible to collect the full spectral trajectory for Batch 15. Therefore, only the results for batch 11, batch 12 and batch 13 are presented. The spectral batch profiles of batch 11 and batch 13 shows again a consistent behaviour as expected. However, batch 12 shows a clear distinctive pattern for all three predicted quality characteristics. The first observation is the deviating behaviour in the early evolution of the batch run (observation 0 – 500). From Figure 4(c) it seems that a certain start-up time was necessary to reach a similar level as batch 11 and batch 13. Further analysis is necessary to search for a cause of this behaviour. Moreover, the relation between the changes in the logged process variables and the accompanying changes in the spectral profiles should be explored.

Taking into account the information given in the previous section and, therefore, ignoring batch 12, it can be concluded that the two remaining spectral batch profiles follow a similar pattern implying a consistent operation.

An additional note should be made for the estimated profiles of the attributes. As can be seen from the time series for the predicted attributes, there is a strong correlation between the predicted attributes. At this point it is not clear if the models are able to predict the individual attributes or a combination of the attributes. That is, is the model predicting a correlation of two attributes or are the models able to discriminate between the individual attributes. To confirm that the actual attributes are predicted additional experiments are carried out.

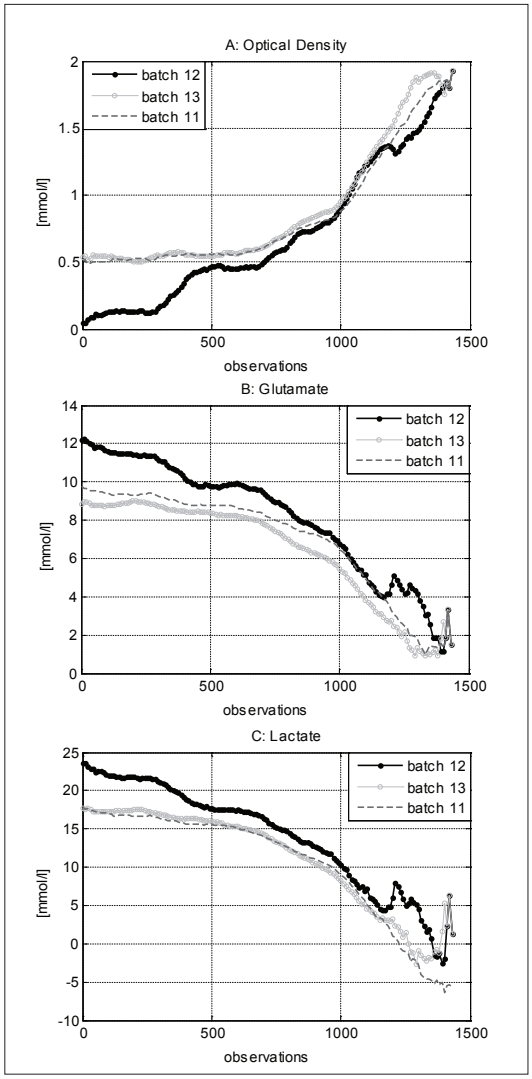


Figure 4. Predicted values for optical density (a), glutamate (b) and the lactate (c) during the batch operation.

Sensor fusion: combining single channel sensors with multi-channel sensors

Studying the responses of the different types of sensors separately gives complementary information about the process. Sometimes it is also revealing to combine the sensors in one global analysis; this is called the sensor fusion approach. A few suggestions on how to

combine the two types of sensors will be discussed shortly.

Depending on the situation, there are different strategies to follow. If it is not possible to develop calibration models for the attributes of interest, a performance model can be developed taking into account the different nature of sensors. This can be done by constructing a concatenated observation matrix $\mathbf{X}_{obs}$,

$$X_{obs} = \left[X_{\text{process variables}} \quad X_{\text{spectra}} \right], \tag{12}$$

where $\mathbf{X}_{process\ variables}$ are the logged process variables and $\mathbf{X}_{spectra}$ the measured NIR spectra. By taking into account the nature of the data and a subsequent proper pre-processing of the data, a model can be constructed on this concatenated matrix. Using the model in a similar way as discussed previously in Section 4.2, an overall view of the batch behaviour will be obtained. That is, because of the nature of the used modelling approach, spectral information is merged with the engineering variables[19,20,21].

If a distinction has to be made between the absorbing components and engineering data, it is possible to summarise the spectra by the predictions of the attributes of interest. That is, if it is possible to develop calibration models for the attributes of interest, the predictions can be used to develop a new model. This is done by taking the predictions of the *Q* attributes and arrange these in a matrix $\mathbf{A}_{con}$ (*K* x Q) and simply construct an observation matrix made up of the logged process variables $\mathbf{X}_{proces\ variables}$ and $\mathbf{A}_{con}$ (*K* x Q).

$$X_{obs} = \left[X_{\text{process variables}} \quad A_{\text{con}} \right] \tag{13}$$

Next, a model is developed by applying a similar approach as above (Section 4.2) and the operational consistency is evaluated. In addition, if deviations occur it is now much easier to assign causes to this deviation. With the current matrix $\mathbf{X}_{obs}$ the individual contribution of each of the variables to the observed signal is assigned by means of contribution plots but that is beyond the scope of this paper[22].

Conclusions

In this paper, PAT related data are presented obtained during the execution of the batch-wise cultivation of *Bordetella pertussis*. In addition, well-defined seed material and a completely defined medium have been used. These are ideal production conditions to illustrate the use of two characteristic tools mentioned in the PAT guidance document for the most critical part of the manufacturing of biopharmaceuticals.

The results are presented of multivariate analyses obtained of process data and data obtained from the NIR analyzer. The results showed that three of the four cultivations are comparable in terms of evolutionary behavior. The NIR data are presented in a quantitative way (optical density and the two substrate levels) and showed a somewhat less reproducible picture than expected.

Based upon the information obtained in this paper, the conclusion can be drawn that the reproducibility of the cultivation step is high and that both types of PAT tools are very appropriate to monitor the upstream part of manufacturing processes of biopharmaceuticals.

We realize that this paper is only the first step in the introduction of the two PAT tools for the manufacturing of biopharmaceuticals. Despite the limitations due to the size of the study, the data presented are supporting the feasibility of the PAT initiative of the FDA for the manufacturing of biopharmaceuticals.

Acknowledgement

The authors want to thank the reviewers for providing comments on the manuscript.

References

1. Streefland, M., Van de Waterbeemd, B., Happe, H., Van der Pol, L. A., Beuvery, E. C., Tramper, J., and Martens, D. (2007) PAT for vaccines: The first stage of PAT implementation for development of a well-defined whole cell vacinne against whooping cough disease. *Vaccine, in press*.
2. Thalen, M., van den IJssel, J., Jiskoot, W., Zomer, B., Roholl, P., de Gooijer, C., Beuvery, C., and Tramper, J. (1999) Rational medium design for Bordetella pertussis: basic metabolism. *J.Biotechnol.* 75(2-3):147-159.
3. Jackson, J. Edward, (2003) A users guide to principal components, Wiley, New York.
4. Nomikos, P. and MacGregor, J. F. (1994) Monitoring Batch Processes Using Multiway Principal Component Analysis. *Aiche Journal* 40(8):1361-1375.
5. Nomikos, P. and MacGregor, J. F. (1995) Multivariate Spc Charts for Monitoring Batch Processes. *Technometrics* 37(1):41-59.
6. Wold, S., Kettaneh, N., Friden, H., and Holmberg, A. (1998) Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemometrics and Intelligent Laboratory Systems* 44(1-2):331-340.
7. Lee, D. S. and Vanrolleghem, P. A. (2003) Monitoring of a sequencing batch reactor using adaptive multiblock principal component analysis. *Biotechnology and Bioengineering* 82(4):489-497.
8. Van Sprang, E. N. M., Ramaker, H. J., Westerhuis, J. A., Gurden, S. P., and Smilde, A. K. (2002) Critical evaluation of approaches for on-line batch process monitoring. *Chemical Engineering Science* 57(18):3979-3991. 10.
9. Kourti, T. (2006) Process analytical technology beyond real-time analyzers: The role of multivariate analysis. *Critical Reviews in Analytical Chemistry* 36(3-4):257-278.
10. Lopes, J. A., Costa, P. F., Alves, T. P., and Menezes, J. C. (2004) Chemometrics in bioprocess engineering: process analytical technology (PAT) applications. *Chemometrics and Intelligent Laboratory Systems* 74(2):269-275.
11. Wold, S., Cheney, J., Kettaneh, N., and McCready, C. (2006) The chernometric analysis of point and dynamic data in pharmaceutical and biotech production (PAT) - some objectives and approaches. *Chemometrics and Intelligent Laboratory Systems* 84(1-2):159-163.
12. Ramaker, H. J., van Sprang, E. N. M., Westerhuis, J. A., and Smilde, A. K. (2003) Dynamic time warping of spectroscopic BATCH data. *Analytica Chimica Acta* 498(1-2):133-153.
13. Small, G. W. (2006) Chemometrics and near-infrared spectroscopy: Avoiding the pitfalls. *Trac-Trends in Analytical Chemistry* 25(11):1057-1066.
14. Gurden, S. P., Westerhuis, J. A., and Smilde, A. K. (2002) Monitoring of batch processes using spectroscopy. *AIChE Journal* 48(10):2283-2297.
15. Ramaker, H. J., van Sprang, E. N. M., Gurden, S. P., Westerhuis, J. A., and Smilde, A. K. (2002) Improved monitoring of batch processes by incorporating external information. *Journal of Process Control* 12(4):569-576.
16. Van Sprang, E. N. M., Ramaker, H. J., Westerhuis, J. A., Smilde, A. K., Gurden, S. P., and Wienke, D. (2003) Near-infrared spectroscopic monitoring of a series of industrial batch processes using a bilinear grey model. *Applied Spectroscopy* 57(8):1007-1019.
17. Olivieri, A. C., Faber, N. K. M., Ferre, J., Boque, R., Kalivas, J. H., and Mark, H. (2006) Uncertainty estimation and figures of merit for multivariate calibration. *Pure and Applied Chemistry* 78(3):633-661.
18. Wold, S., Sjostrom, M., and Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58(2):109-130.
19. Puxty, G., Maeder, M., Rhinehart, R. R., Alam, S., Moore, S., and Gemperline, P. J. (2005) Modeling of batch

reactions with in situ spectroscopic measurements and calorimetry. *Journal of Chemometrics* 19(5-7):329-340.

20. Smilde, A. K., Westerhuis, J. A., and Boque, R. (2000) Multiway multiblock component and covariates regression models. *Journal of Chemometrics* 14(3):301-331.
21. Smilde, A. K. and Kiers, H. A. L. (1999) Multiway covariates regression models. *Journal of Chemometrics* 13(1):31-48.
22. Westerhuis, J. A., Gurden, S. P., and Smilde, A. K. (2000) Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems* 51(1):95-114.

CHAPTER 6

Exploration and Modelling of the Design Space of a Bacterial Vaccine Cultivation Process

Streefland M[1*], Van Herpen PFG[1], Van de Waterbeemd B[1], Van der Pol LA[1], Beuvery EC[2], Tramper J[3], Martens DE[3], Toft M[4]

[1] Netherlands Vaccine Institute, Unit Research and Development, PO Box 457, 3720 AL Bilthoven, The Netherlands
[2] PAT Consultancy, Kerkstraat 66, Vianen, The Netherlands
[3] Wageningen University, Department Bioprocess Engineering, Bomenweg 2, 6781 EV Wageningen
[4] Umetrics AB, Stortorget 21, 211 34 Malmö, Sweden

Submitted to Biotechnology and Bioengineering

Abstract

A licensed pharmaceutical process is required to be executed within the validated ranges throughout the lifetime of product manufacturing. Changes to the process, especially for processes involving biological products, usually require the manufacturer to demonstrate that the safety and efficacy of the product remains unchanged by new or additional clinical testing.

Recent changes in the regulations for pharmaceutical processing allow broader ranges of process settings to be submitted for regulatory approval, the so called process design space, which means that a manufacturer can optimize his process within the submitted ranges after the product has entered the market, which allows flexible processes.

In this paper, the applicability of this concept of the process design space is investigated for the cultivation process step for a vaccine against whooping cough disease. An experimental design (DoE) is applied to investigate the ranges of critical process parameters that still result in a product that meets specifications. The on-line process data, including near infrared spectroscopy, are used to build a descriptive model of the processes used in the experimental design. Finally, the data of all processes are integrated in a multivariate batch monitoring model that represents the investigated process design space. This paper demonstrates that the general principles of PAT and process design space can be applied to an undefined biological product such as a whole cell vaccine. The approach chosen for model development described here, allows on line monitoring and control of cultivation batches in order to assure in real time that a process is running within the process design space.

Introduction

"The process design space is the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality"[1]. This is the definition of process design space proposed by ICH, the International Conference for the Harmonization of Pharmaceutical Regulation. The word *design* indicates the requirement for a rationale behind process development. The design of a process that can consistently assure the yield of products with the desired quality is one of the key elements of the Process Analytical Technology (PAT) initiative introduced by the American Food and Drug Administration (FDA) in 2004[2]. The FDA defines PAT as a "*system for designing, analyzing and controlling manufacturing through timely measurements of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality."*

The development and use of a design space is concisely described in the Annex to the ICH Q8 Guideline on Pharmaceutical Development[1]. Originally, the concept of PAT and design space applied only to chemical drugs. However, the principles behind PAT and the design space can be readily extended to biopharmaceuticals as is demonstrated in this paper and reported earlier by others[3].

Prior to the introduction of the design space concept, process validation relied initially on the execution of three consecutive conformity batches that demonstrated process robustness. Although this "rule of three" was never mentioned in the official FDA guidance on process validation[4], it became the standard for (bio)process validation. Amendments to the validation guidance eventually resulted in replacement of this rule with the requirement that the manufacturer should demonstrate process reproducibility and consistency throughout full scale manufacturing. In all cases, excursions from the validated process must be reported as *deviation* to the regulatory authorities. Thus, manufacturers are bound to execute their processes within the validated specifications, unless they revalidate the changed process with additional product testing that demonstrates *product comparability*[5]. Sometimes even additional clinical trials are required. For complex biopharmaceuticals such as classical vaccines, this "process is the product" paradigm made it almost impossible to make changes to the process without clinical confirmation that the product remains potent, safe and otherwise comparable to the original product.

The design space concept allows manufacturers to move away from fixed processes, permitting process variation after sufficient understanding and control of the process and knowledge of the impact of process variation on product performance have been demonstrated. In such a case, a change of the process settings or trajectory within an approved design space is not considered a deviation in the regulatory sense of the word (i.e. it does not require regulatory approval).

One way of demonstrating the understanding of the process and its variation is to investigate

the critical process parameters using Design of Experiments or DoE[6]. DoE investigates not only the impact of the variation of a single, critical parameter, but also any interactions that these critical parameters might have with each other. For instance, a process might be running fine at high temperature or low pH, but the combination of these two process conditions could be unfavorable for product quality.

The mapping of a process begins with an understanding of the critical product quality attributes, for example, in the case of vaccines, the antigens that induce the protective immune response. The next step is the identification of those process parameters that influence process performance (e.g. yield) or the critical product quality attributes (e.g. antigen expression). These parameters can be anything from the quality of (complex) starting materials (such as bovine serum or yeast extract) to the pH and temperature during processing.

Once the critical process and product attributes are known, the failure limits of the process can be investigated. These are the limits of critical process parameters at which the critical product quality attributes are so compromised that the product will no longer meet its specifications or that the process will fail altogether. With existing products, some specific limits are well known to the operators that run the process routinely, while others are generally known to be common for a broad range of processes. For instance the upper temperature limit of a cultivation process step can usually not exceed 40 °C for most organisms, because cellular processes are radically changed or even disrupted at high temperatures. The failure limits or critical limits for any remaining parameters can be investigated with a series of simple experiments, such as we reported for dissolved oxygen concentration[7].

The combination of the failure limits does not, by itself, constitute a design space since it provides no means for predictions outside of the actually tested settings. Rather, the knowledge gained from the evaluation of failure limits can be used to devise a DoE experimental matrix to determine the design space. The results from a DoE experimental matrix can be modelled so that predictions can be made for a continuous range of settings and their interactions, based on a discrete number of tested settings. Test values for the critical process parameters that are known to yield product of specified quality within the process failure limits are selected for the matrix. The tested ranges of the critical parameters then provide the boundaries of a multidimensional space having the number of its dimensions equal to the number of critical parameters in the design.

Biopharmaceutical products are typically produced in a batch-wise manner, and this gives rise to three-way data arrays: batch number vs. processing time vs. process parameters (see also Fig. 2). Chemometric methods such as principal component analysis (PCA) and partial least squares (PLS) calibration models[8] can be used to model the three-way batch data in order to provide tools for understanding, fault detection, control, and prediction. The resulting dynamic control models can be used to describe a (partial) design space, taking into account the evolution of parameters during processing.

In this paper, we investigate the design space for the cultivation of *Bordetella pertussis* bacteria. This is the key process step in the manufacturing of a whole cell vaccine against whooping cough disease. A whole cell vaccine means that the whole bacterial cell is the actual product, in contrast to, for instance, sub unit vaccines in which only purified membrane proteins are used. The downstream processing of such a vaccine involves only inactivation and concentration of the cells without a purification step. Thus, the most critical product quality characteristics are already defined at the end of the cultivation process.

The DoE explores the design space for this process step through variation of the critical independent process parameters (within their failure limits) in order to determine what, if any, combination of parameter values results in insufficient product quality. The failure limits were usually known before the experimentation, this knowledge resting on the fact that the Netherlands Vaccine Institute has 40 years of manufacturing experience with this product. The purpose of the investigation was to demonstrate that the combination of tested ranges for all critical process parameters still results in good product quality. In this case, the entire process space under investigation constitutes the design space.

This approach may be contrasted with that for new products. In the that case, larger ranges of values of the critical process parameters are initially investigated, and the *design space* is determined as a sub-space of the *investigated space* (sometimes referred to as k*nowledge space*, Figure 1). Such an approach is necessary when there is little or no manufacturing experience and insufficient knowledge of the process failure limits.

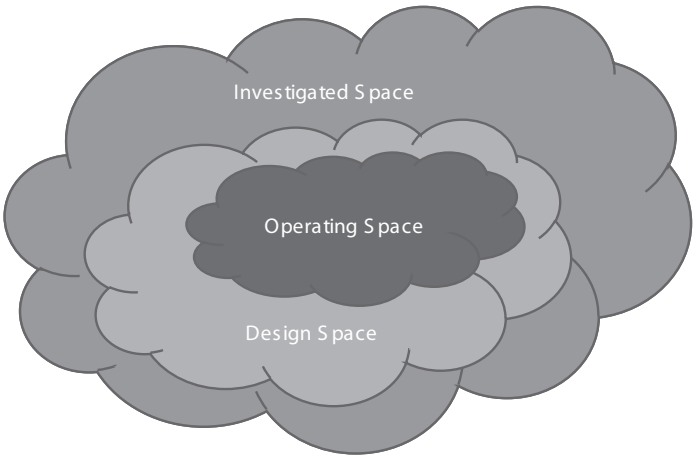


Figure 1. The Design Space as a sub-space of the Investigated Space (also known as the Knowledge Space). The Operating Space (where the process is intended to run) can be a large or small part of the Design Space. (Adapted from (Harms et al. 2008).

In order to *consistently ensure a predefined quality at the end of the manufacturing process* (ICH definition of the scope of PAT)[1], an additional signal may be needed that can be related to process performance or product quality. According to the FDA's PAT guidance[2], "Measurements collected from these process analyzers need not be absolute values of the attribute of interest". In this study, near infrared (NIR) spectroscopy is investigated as a PAT tool, i.e. *an online measurement of critical quality and performance attributes* (FDA/ ICH PAT definition), to monitor the cultivation of *Bordetella pertussis* bacteria. NIR data are used to indicate specific attributes of interest (such as biomass concentration) and as a "fingerprint" that gives qualitative information on the status and trajectory of the process. All data gathered online – NIR as well as conventional cultivation process data (i.e. pH, temperature, and gasflow controller outputs) - are combined in a single database using PAT compliant software. This database contains all data used to construct the model describing the multivariate process design space.

The design, experimentation and multivariate modelling required to define the design space for the cultivation of *Bordetella pertussis* are described. The resulting process model integrates bioreactor control data with online NIR measurements and can be used for on–line, real time modelling of the cultivation process. This provides real time assurance of process performance and an indication of product quality during processing. Ultimately, the system will be capable of real time release of the cultivation product for further processing.

The main goal of this paper is to demonstrate the principles of design space investigation for a biopharmaceutical cultivation process. We investigate both the experimental work involved, the analytical (PAT) tools used and the data processing and modelling necessary in order to devise a model that describes the process design space. It is a first attempt to apply the principals of PAT and process design space for an undefined biopharmaceutical product, namely a whole cell vaccine. In general, such a product is regarded as the "worst case" in terms of how defined and how well characerized the product is. The ability to apply PAT principles on this product bodes well for the implementation of PAT on other, more defined (bio)pharmaceutical products.

Materials and Methods

Experimental design

An experimental design was executed and data from the design used to mathematically model the effects of the process parameters on product quality. A robustness design in which the parameter values chosen were at a safe margin from known or expected failure limits was employed. The design was a Plackett Burman design in 8 runs to which two times two replicate center points were added (two for each of the two reactors). The experimental matrix, having 12 runs, is shown in Table 1. Experiment 6 was done twice, owing to a data collection problem, once in Reactor A and once in Reactor B.

The process parameters investigated were: the dissolved oxygen (DO), the pH, the temperature, the density of bacteria at the end of the pre-culture phase in the shake-flask (Preculture density), the actual density of bacteria at the start of the bioreactor run (Inoculation density), and which of the two identical bioreactor systems, A or B, was used in the experiment.

Table 1. The DoE matrix for the 12 batch runs used to determine the process design space.

Experiment Number	Run Order	DO (%)	pH	Temperature (°C)	Preculture density	Inoculation density	Reactor
1	1	10	6.8	37	0.5	0.1	A
2	12	60	6.8	33	0.5	0.025	B
3	2	60	6.8	37	1.5	0.025	A
4	3	30	7.2	35	1	0.05	B
5	4	30	7.2	35	1	0.05	A
6B	10	10	7.8	37	0.5	0.025	B
6A	13	10	7.8	37	0.5	0.025	A
7	5	60	7.8	33	0.5	0.1	A
8	6	30	7.2	35	1	0.05	B
9	7	10	7.8	33	1.5	0.025	A
10	8	60	7.8	37	1.5	0.1	B
11	11	30	7.2	35	1	0.05	A
12	9	10	6.8	33	1.5	0.1	B

Bioreactor Cultivations

Cultivation conditions

All cultivations were carried out in a fully instrumented 7 L *in situ*, sterilizable bench-top bioreactor with a six-bladed Rushton turbine (Applikon, Schiedam, The Netherlands). The bioreactor was filled with 4 L of THIJS medium[9] and raised to the set point temperature and a 100% DO condition. It was then inoculated with preculture from a shake-flask at an optical density at 590nm (OD_{590}) of 0.5, 1.0 or 1.5. The volume of preculture used to inoculate the bioreactors was adjusted to yield the desired starting density. Temperature, DO, pH and stirrer speed were controlled at the different values shown in Table 1. A low-drift polarographic electrode (Applikon, Schiedam, The Netherlands) was used to measure the DO concentration in the liquid. pH and temperature were measured using a glass pH electrode (Mettler Toledo, Tiel, The Netherlands) and a Pt100 temperature sensor (Applikon, Schiedam, The Netherlands), respectively. All sensors were connected to the bioreactor

control system (Applikon, Schiedam, The Netherlands), which was operated using PCS7-based bioreactor control software (Siemens, Zwijndrecht, Belgium). DO was controlled with increments in stirrer speed up to 650 rpm after which it was controlled by increasing the fraction of oxygen in the headspace. The fraction of AIR, O_2 and N_2 were also registered in the control system. Total gas flow was kept constant at 1.0 L/min.

Nutrient concentration analysis

Samples were taken at regular intervals and OD_{590} and nutrient concentrations were measured. Samples were sterile filtered (0.22 μm) and supernatants were stored at -20 °C for further nutrient analysis. Lactate and glutamate concentrations were determined by ^{1}H-Nuclear Magnetics Resonance (NMR) spectroscopy using a JEOL JNM ECP 400 spectrometer operating at 400MHz (JEOL, Tokyo, Japan) and equipped with a JEOL Stacman autosampler for 16 samples. Supernatants were analyzed by adding 0.1 ml of D_2O containing 3-(trime thylsilyl)[D4]proprionic acid sodium salt (TSP, 0.167 mM) to a 0.9 ml sample. The water signal was suppressed by a standard pulse experiment using presaturation. The spectra were referenced using the TSP signal at 0 ppm. Lactate and glutamate concentrations were quantified by integration of the relevant signals. NMR was also used to check samples for any waste metabolites.

Data collection and handling

All on line data necessary to build the process model were collected using the SIPAT software. Every two minutes this software collected the data from the cultivation system and the NIR system (see below) and stored it in a central database. When two bioreactor systems were running simultaneously, the SIPAT software collected the data from each system sequentially every minute so that each system was still sampled every two minutes. The SIPAT database was later used to build the process model.

NIR spectroscopy measurements

An on-line NIR transmission probe (Solvias AG, Basel, Switzerland) with a fixed path length of 5 mm, was implemented in both bioreactor vessels. The probe was immersed in the liquid with the measurement slit at the height of the stirrer, pointing away from the stirrer. Measurements were made every two minutes on both probes using a Bruker Matrix F NIR source (Bruker Optics, Ettlingen, Germany). Spectra were taken at a resolution of 4 cm^{-1} between 12000 and 4000 cm^{-1}. Each spectrum was an average of 32 scans.

Additional samples for PLS calibration of biomass density

Some additional samples were necessary to build a suitable PLS calibration model for biomass density. The samples were prepared by cultivating *Bordetella pertussis* in a glass shake-flask according to the standard procedure. At the end of the cultivation (OD_{590}=1.3 – 1.5), half of the shake-flask was centrifuged and sterile filtered to yield cell-free supernatant. Suspension and cell free supernatant were mixed to create samples with optical densities that match, the start, half-time and end of the cultivation, but that have nutrient concentrations that do not match these cultivation stages. In this way, the correlation between optical density and nutrient concentration is disconnected, in order to see if these parameters individually correlate to the NIR signal and that they not merely have the same process trend. The samples were prepared according to Table 2.

Table 2. Sample preparation matrix for extra samples measured off line using NIR.

Sample	Supernatant (ml)	Suspension (ml)	Lac (µl)	Glu (µl)	H2O (µl)
1	1	0	0	0	7.2
2	1	0	0	1.8	5.4
3	1	0	0	3.6	3.6
4	1	0	1.8	0	5.4
5	1	0	1.8	1.8	3.6
6	1	0	1.8	3.6	1.8
7	1	0	3.6	0	3.6
8	1	0	3.6	1.8	1.8
9	1	0	3.6	3.6	0
10	0.5	0.5	0	0	7.2
11	0.5	0.5	0	1.8	5.4
12	0.5	0.5	0	3.6	3.6
13	0.5	0.5	1.8	0	5.4
14	0.5	0.5	1.8	1.8	3.6
15	0.5	0.5	1.8	3.6	1.8
16	0.5	0.5	3.6	0	3.6
17	0.5	0.5	3.6	1.8	1.8
18	0.5	0.5	3.6	3.6	0
19	0	1	0	0	7.2
20	0	1	0	1.8	5.4
21	0	1	0	3.6	3.6
22	0	1	1.8	0	5.4
23	0	1	1.8	1.8	3.6
24	0	1	1.8	3.6	1.8
25	0	1	3.6	0	3.6
26	0	1	3.6	1.8	1.8
27	0	1	3.6	3.6	0

The lactate and glutamate stock solutions were prepared at 4.44 and 2.22 M, respectively. The amounts listed in the table above produced concentrations typical for the start, halftime or end of a standard cultivation. The half-time samples were prepared by mixing cell-free supernatant with cell suspension at a dilution of 1:1. The cell-free supernatant was used as a surrogate for the low-OD_{590} samples.
Samples 1-27 were prepared according to the table and, after thorough mixing, were quickly measured using a NIR cuvette bench (Bruker Optics, Ettlingen, Germany). The settings were identical to those used for the transmission probe inside the bioreactor.

Microarray Analysis

RNA isolation

Samples for microarray analysis were taken at the end of each cultivation. For fixation of the RNA expression profile, one volume of bacterial culture was mixed with two volumes of *RNase retarding solution*[10,11,12]. For each microarray sample 2.5 ml cultivation at OD_{590} = 1.0 was used. Samples at other optical densities were adjusted accordingly, so that an equal amount of cells was used for each sample. The samples were concentrated by centrifugation and treated for 3 minutes with Tris-EDTA buffer, containing 0.5 mg/ml lysozyme (Sigma-Aldrich, Zwijndrecht, The Netherlands). Total RNA was extracted with the SV total RNA isolation system (Promega Benelux, Leiden, The Netherlands) according to the manufacturer's protocol. Nucleic acid concentration was adjusted by precipitation. UV spectral analysis was used to determine final nucleic acid concentration and purity and RNA integrity was confirmed with the Bioanalyzer RNA6000 Nano assay (Agilent Technologies, Amstelveen, The Netherlands), according to the manufacturer's protocol.

cDNA labeling and hybridization reactions

Total RNA from all experimental samples was reverse transcribed to cDNA and labeled with Cy3/Cy5 dyes using the Chipshot Indirect Labeling Kit (Promega Benelux, Leiden, The Netherlands) according to manufacturer's protocol, with one deviation: 2 μl random nonamer primer and no oligo-dT primer was used per reaction to reverse transcribe the total RNA, because prokaryote mRNA does not have poly-A tails. Experimental samples (Cy5) were pooled with a common reference sample (Cy3) having equal amounts of RNA from all experimental samples in the experiment. Volumes of the combined cDNA samples were adjusted to 25 μl and an equal amount of hybridization buffer was added, to a final concentration of 25% formamide, 5x SSC and 0.1% SDS. Samples were applied to the microarray slides and placed in a hybridization chamber (GeneMachines, San Carlos, CA, U.S.A.) for 16-20 hours at 42° C in the dark.
The microarrays were scanned with a ScanArray Express microarray scanner (Perkin Elmer,

Groningen, The Netherlands) and median fluorescence intensities were quantified for each spot using ArrayVision software (Imaging Research, Roosendaal, The Netherlands).

Processing of the expression data

The expression data were natural-log transformed, quantile normalized, corrected for the common reference dye signal and the values of replicate spots were averaged. These data processing steps were carried out using the free statistical software R, running an in-house developed script.

A product quality score was calculated from the gene expression data as described earlier (see Chapter 4). To calculate the product quality score, only genes from the *virulence core regulon*[13,14] were considered, since these are the genes that are considered to be critical for the function of the vaccine. Based on earlier work, a product quality score of 8 or higher, can be considered to represent good quality (see Chapter 4).

Data Analysis and Modelling

The data available in this study can be categorized as follows (Fig 2):

- X_{DoE}: the setup parameters from the DoE (T, pH, DO, preculture density, inoculation density and reactor)
- X_{proc}: the time varying process variables measured throughout the batch evolution (pH, DO, T, flow of air, O_2, and N_2, stirrer speed)
- X_{NIR}: the on-line NIR data measured throughout the batch evolution
- Y_{procQ}: the process variables related to process evolution measured at-line by taking out samples during the cultivation (OD_{540}, glutamate and lactate)
- Y_{prodQ}: the final product quality variables measured after the completion of the batch, i.e. the microarray expression data (product quality score, see paragraph 3.3)

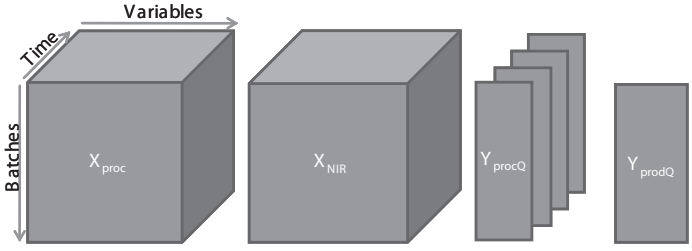


Figure 2. Overview of available data blocks.

Due to the large number of distinct data sets available for the study, the analysis was split up into four sub-steps.

Step 1: Traditional design space analysis

The gene expression data (Y_{prodQ}) were transformed into a product quality score and the relationship between the design setup parameters (X_{DoE}, Table 1) and the product quality score was modeled to detect significant effects. The approach involved a traditional DoE analysis in which the output of the experiments was associated with the input settings. Multiple linear regression (MLR) was used to model the relationship between the process parameters and the product quality using the MODDE (Umetrics AB, Malmö, Sweden) DoE software package. A robust design space should result in a non-significant model in which no significant relationship exists between product quality and variation in the process setting within the tested range. In that case the process is stable within the design space.

Step 2: Prediction of process variables from NIR data

The process evolution variables (Y_{procQ}) OD_{590}, lactate and glutamate are a measure of the status of the cultivation and can be used to assess changes in specific growth rate or metabolism. However, these parameters are only measured at distinct sample points during cultivation. It is of interest to have these variables available on line for more accurate monitoring of the status of the cultivation. NIR data was used in an attempt to predict these process evolution variables. NIR data values were used as an X data matrix with the process evolution variables as the Y data matrix. A PLS regression model was fitted to the two datasets. The NIR data were centered before fitting the PLS model, while the process evolution data were centered and scaled to unit variance.

Initial modelling revealed a high degree of correlation between the three process evolution variables, OD_{590}, lactate and glutamate. Therefore, a prediction model based only on the samples taken from the cultivations was not reliable. Additional samples having the three evolution variables uncorrelated with each other were therefore prepared as described in Table 2. These samples were also subjected to NIR spectroscopy analysis and the data were added to that from the cultivations to obtain a reliable prediction model.

Step 3: PCA fingerprints from NIR data

The NIR measurement results in more than 4000 variables, which are not mathematically independent and which do not all carry unique information. To use NIR data as a "fingerprint" of the status of the process, the spectra are compressed into a small number of independent Principal Components that summarize the variation in the large number of initial variables. The computational time needed for the calculation was reduced by using one spectrum per hour for the calculation of the PCA model for each batch. Finally, all spectra (taken every 2 minutes) were compressed using this model.

The "fingerprint" derived from the NIR spectra doesn't necessarily represent specific physical or chemical characteristics of the samples, but may contain more abstract characteristics. It is assumed, however, that the NIR fingerprints are similar for all batches, thus that it can be used as a non-specific variable for the monitoring of batch evolution, allowing non-biased batch-to-batch comparisons.

Step 4: Building the process control model

The process data registered during batch evolution (denoted X_{proc} in Figure 2) were modeled together with the OD_{590} predicted from NIR (Step 2) and the NIR fingerprint variables (Step 3). Together they constitute a three-dimensional data matrix of batches vs. variables (process variables, predicted OD_{590} data and NIR fingerprint variables) vs. process time.

The evolution was split up into two phases for each batch. The first phase was represented by the part of the cultivation in which Dissolved Oxygen concentration (DO) was controlled by increments in stirrer speed and the second in which DO was controlled by increasing the fraction of O_2 in the headspace. This was done to allow for a different correlation structure in the process variables before and after the change in DO control strategy.

The data were modeled as described in Wold et al.[8] using the modelling setup denoted Observation-Wise Unfolding with subsequent Batch-Wise Unfolding of the scores (OWU-BWU).

Software

ArrayVision 8.0 (GE Healthcare, Uppsala, Sweden) was used to quantify microarray fluorescence intensities.

R (WU Wien, Austria) was used for statistical analysis of gene expression data.

MODDE (Umetrics, Umeå, Sweden) version 8 was used for setting up the design and for the traditional DoE analysis.

SIMCA-P+ version 11.5 (Umetrics, Umeå, Sweden) was used for all NIR data analysis as well as for building the process control model.

The SIPAT software (Siemens, Nynove, Belgium) was used for collecting process and NIR data during the cultivations and served as the database for extracting data for use in the NIR and process models.

Results and Discussion

Traditional Design Space Analysis

The product quality at the end of each of the 13 runs was determined by calculating the product quality score for each batch (see Chapter 4) from the gene expression data (Table 3). Prior experimentation showed that good quality is represented by a score above 8 (see Chapter 4), and thus the scores in Table 3 (all higher than 8.5) should represent good product quality.

Table 3. DoE experiments and their corresponding product quality scores

Experiment number	Product Quality Score
01	10,36
02	10,00
03	9,95
04	8,89
05	9,68
06B	9,74
06A	8,57
07	9,67
08	10,07
09	9,51
10	9,91
11	10,58
12	10,53

The quality scores are used instead of the actual animal (Kendrick) tests for product quality that are normally required by the authorities for whooping cough vaccine. This test involves an intra-cranial challenge of mice previously vaccinated with a human dose of vaccine with live *Bordetella pertussis* bacteria[15,16]. This test is not only highly variable and very cruel to the animals, but also costly and labor intensive[17]. The product quality scores are therefore better suited for building this preliminary process model.

Using the DoE setup parameters as input (X) variables and the product quality scores as output (Y) variables, a traditional DoE analysis was done using MODDE software. No significant model could be fitted, (best fit: $R^2 = 0.38$; $Q^2 = -0.2$) indicating that, within the ranges of the variables of the DoE setup, variation in the value of the input parameters had no effect on the product quality. This indicates that, within the tested ranges, the process is robust. The ranges tested in this study can therefore qualify as a design space as intended in the ICH Q8 Guideline[1]. The data collected during these experiments can be used to fit a model that describes the process and can later be used as an on-line check of whether a new

process is running within the design space.

Prediction of Process Variables from NIR Data

A raw data plot of the NIR data is shown in Figure 3. The plot shows clear offset differences between the spectra, probably due to difference in optical density of the samples. No attempt was made to remove this variation as this parameter is probably the main correlate for optical density in the NIR signal, which is one of the process evolution variables to be predicted from the NIR spectra.

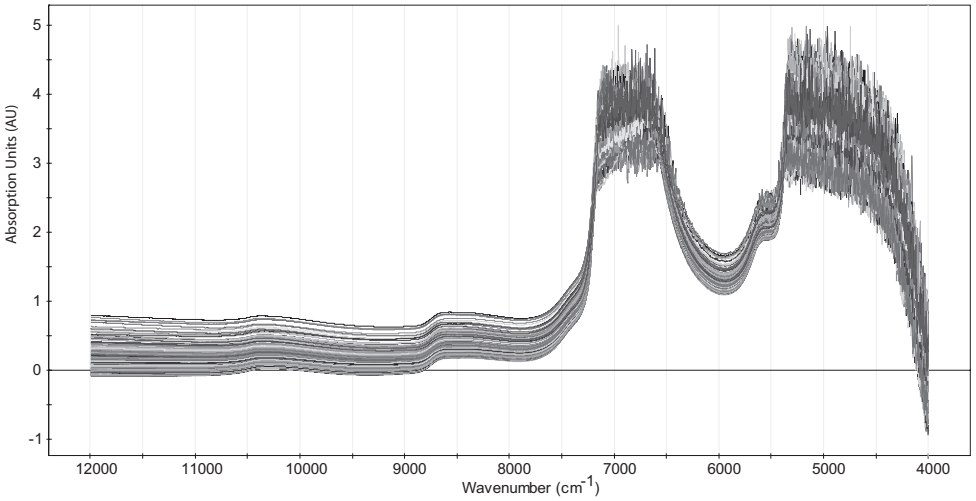


Figure 3. Raw data plot of NIR spectra of the samples taken out during cultivation.

The very strong signals from water in the spectral regions 7200 – 6500 and 5400 – 4000 cm^{-1} saturated the detector and, therefore, models built on the NIR spectra exclude these regions. The benefit of these high sensitivity measurements (too high for the water peaks) is that the signals in the other regions are stronger.

A PLS model was built from the NIR spectra (X) and all three process evolution variables: OD_{590}, lactate and glutamate (Y). This model provided very good results with a cross-validated explained variance (Q^2_Y) of 81-97% for the three parameters. However, the three parameters were highly correlated, because OD_{590} is initially low in a cultivation, while the concentration of nutrients is high. As the cultivation proceeds, the OD_{590} increases while concentration of nutrients decrease at the same rate. Thus predictions from this model cannot be trusted, because the model is unable to predict samples correctly where this relationship between the three variables is not present. Artificial samples were prepared (Table 2) in order to break up the correlation between the three process evolution variables. PLS models built

only on these samples showed that OD_{590} could be reasonably predicted, while no significant components were found for the prediction models for lactate and glutamate.

To incorporate any difference between the spectra for the artificial samples prepared from a shake-flask cultivation and the samples taken directly from the bioreactor cultivations, a PLS model for predicting OD_{590} that was based on both sets of samples was built (total samples number: 85; 58 original cultivation samples and 27 additional samples). The model showed very good predictive ability (Q^2_Y was 96%) and no difference between prediction error of the artificial samples and the original cultivation samples was seen (Figure 4). The predicted values of OD_{590} were used in the subsequent analysis of batch evolution together with the process variables. This prediction model, based on the NIR signal, can be executed during cultivation of future batches, allowing a growth curve to be plotted in real time without the need for excessive sampling. However, the prediction of the higher values OD_{590} still show a large prediction error, which means that more samples from more batches need to be added to the model in order to increase it's reliability before it can replace manual sampling completely.

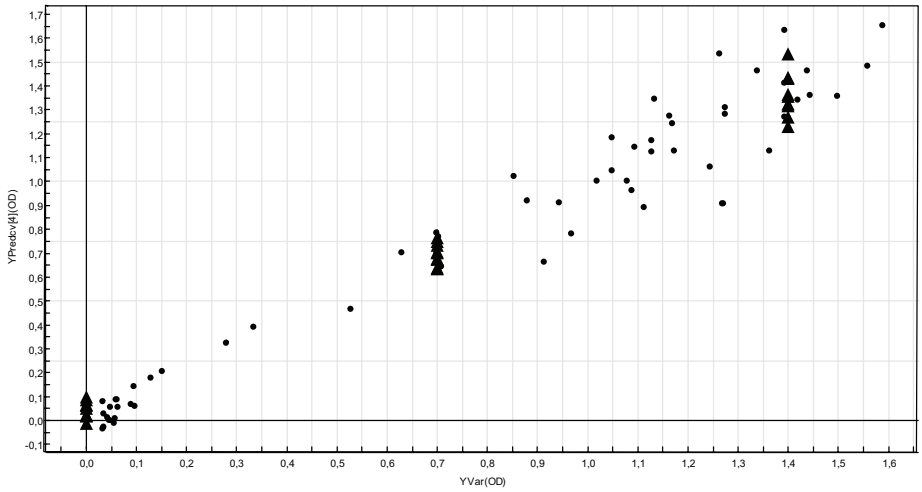


Figure 4. Plot showing values of OD_{590} predicted during cross-validation (ordinate) versus measured/prepared values (abcissa) of OD_{590}. Circles represent samples from original cultivation, while triangles represent artificial samples.

PCA Fingerprints from NIR Data

NIR data can also serve as a fingerprint of the status of the process. For this purpose, a Principal Component Analysis (PCA) based model was created from the NIR cultivations data. To reduce calculation time, one spectrum per hour was used to construct the model (incorporating the entire biological evolution of the batches), resulting in a total of 218 NIR

spectra. The regions containing the water signals were excluded as described above.
A PCA model with five components was created that explained 99.9 % of the variation in the data. All NIR spectra (acquired every 2 minutes throughout the cultivations) were then compressed into these five components using this model. The five components, used as five fingerprint variables (NIR_fp1 to NIR_fp5), together with the process variables are used in the subsequent analysis of batch evolution.

Building the Process Control Model

Integrity of process data

Due to an overnight data collection error, the middle part of the process and NIR data for Batches 3 and 6B were lost (~ 30-40% of total batch data), which means these data could not be used in the model. The data from samples taken from these batches were used, however. Process and NIR data for Batches 2, 8, 10, 11 and 12 were not sampled consistently every 2 minutes throughout batch evolution due to a wrong setting of the NIR sampling method. The interval for these batches varied generally between 1 to 5 minutes, which is adequate for batch modelling, so the batches could be used. The available data were sufficient to build a robust process model. In order to allow all datasets to be processed in a similar way, any gaps in the datasets were filled using averaged values in such a way that they did not influence the model.

Modelling of dynamic evolution

Process variables together with predicted OD_{590} and NIR fingerprint variables were modeled using PLS with time as the Y variable, as described by Wold et al. (8). Prior to modelling, all variables were investigated individually and trimmed to remove unrepresentative values (e.g. spikes caused by air bubbles in the nIR measurement slit). The resulting PLS model based on all on-line signals from the bioreactor control system, the predicted biomass and the NIR fingerprint variables, describes the evolution of the cultivation batches as a result of the settings chosen in the DoE matrix (Table 1). This is called a dynamic evolution model.
This model consists of two phases, each containing four components, which explain >99% of process variance, as shown in Figures 5 and 6. The line graphs show the evolution of the loading plots for each batch over time, which the bars (loading plot) indicate the contribution of each parameter to the observed variance. A detailed interpretation of these components is provided in Appendix 1, while a short summary is given below. Note that it is not critical that all components be fully understood for a monitoring model. When, however, this model is used for the next level of modelling, the multivariate batch monitoring model, it is important that it encompasses “normal” batch evolution, in other words that it describes the process design space completely. It should thus distinguish normal or good batches from deviating batches. Because of the limited number of batches used in the experimental design, this will initially give ambiguous results in parts of the design space that are not

fully covered in the experimentation. Data from new cultivations (for instance from regular manufacturing) should be added to the process model to make it more reliable over time. The model may also be challenged with experimental faulty batches for validation purposes.

The batch evolution was split into two phases coinciding with a switch in the oxygen control strategy (Figure 5 and 6, respectively). At the start of a batch, before inoculation, dissolved oxygen (DO) is maintained at set-point using AIR and N_2 with a total flow of 1 L/min. After inoculation, while maintaining a total flow of 1 L/min oxygen, consumption rises and the fraction of nitrogen decreases to zero. Simultaneously, the stirrer speed is gradually increased to a maximal set-point value. When both AIR and stirrer speed have reached their maximum set-points, the controller switches to a new strategy in which the fraction of oxygen in the gas mixture is increased by mixing pure O_2 into the airflow, while still maintaining total gas flow at 1 L/min. This relationship is seen in the first component for the first phase, where PredictedOD, AIR, stirring, and NIR_fp1 are increasing with batch evolution while N_2 is decreasing.

The second phase of the process (Figure 6) begins when the stirrer speed is no longer increasing and the pure oxygen valve is opened for the first time. This change in control strategy is clearly seen in the first component for the second phase, where PredictedOD and NIR_fp1 continue to increase, while AIR decreases and O_2 is increases. Further, more detailed, explanation of the model components is given in Appendix 1.

Multivariate batch monitoring model

For on-line monitoring whether any given process is running inside the process design space, a multivariate batch monitoring model is required. The batch monitoring model was constructed based on the dynamic evolution model as described in Wold et al.[8]. This model gives a total overview of the difference between the batches. To illustrate this, a PCA model with two components was built on the scores from the dynamic evolution model. The scores plot of this model is shown in Figure 7.

This plot is a 2D graphical representation of the process design space as it is described by the DoE batches. All batches are condensed into one point and the outer circle represents the 95% confidence interval of the (design) space defined by these points. The numbers correspond with the batch numbers shown in Table 1. Batches 3 and 6B are not represented because of insufficient nIR and process data, as explained earlier.

All relevant process data, including the predicted OD and the NIR fingerprint variables, are condensed in the multivariate batch monitoring model. Figure 7 shows that the design applied can be seen to have caused some variation in the process evolution. The batches are similar – none fall outside of the Hotellings T2 ellipse that represents a 95% confidence interval.

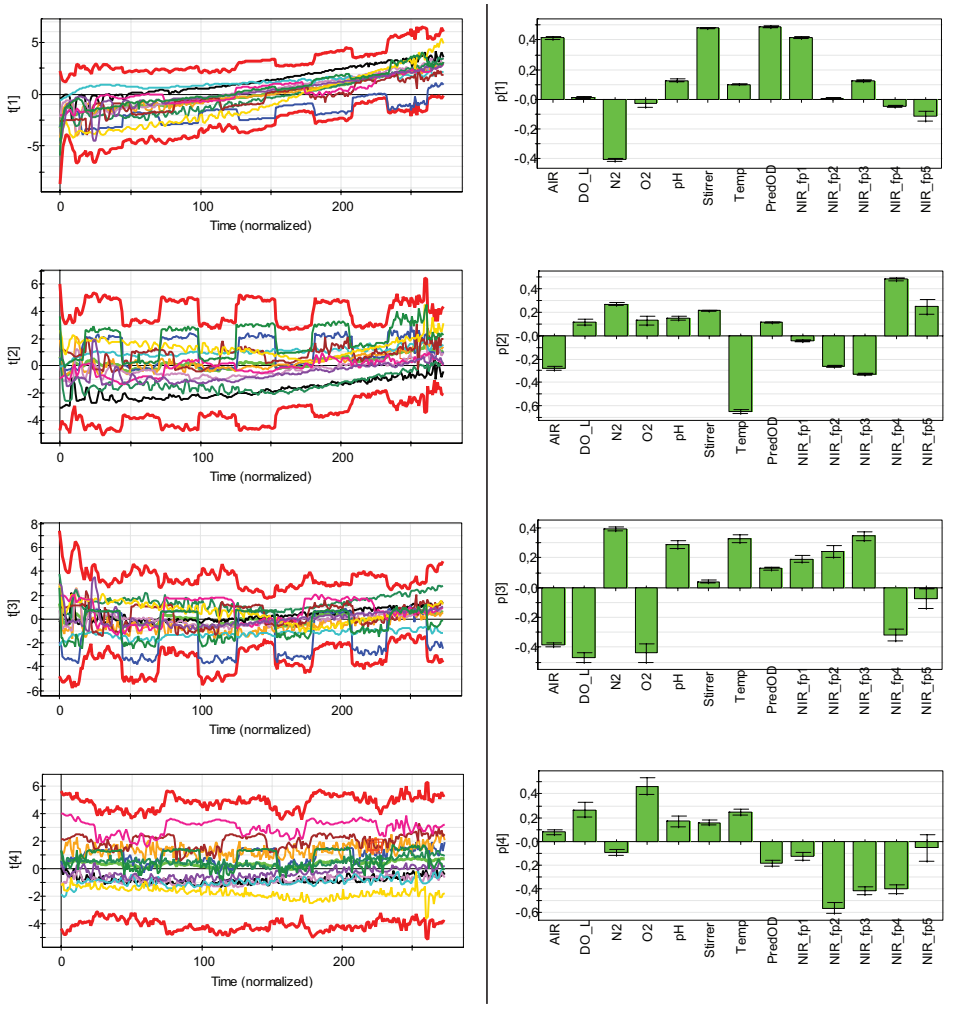


Figure 5. Scores (left) and loadings (right) of Phase 1 of components 1 to 4 (top to bottom) of the dynamic evolution model. The outer red lines in the line graphs indicate a 95% confidence interval and the green line the average. The other lines each represent the evolution of the score of a cultivation batch over time. The loadings (right) are represented as contribution plot in which the contribution of each individual parameter to the variance is shown. Error bars indicate the standard deviation over time.

The design is clearly reflected in the plot, as the centre-points of the design (those batches run with intermediate values of all six investigated parameters), batches 4, 5, 8 and 11, are close together and near the centre of the graph. Furthermore a separation in temperature kan be observed, with at the bottom and right part the cultures at 33 °C (squares), in the middle

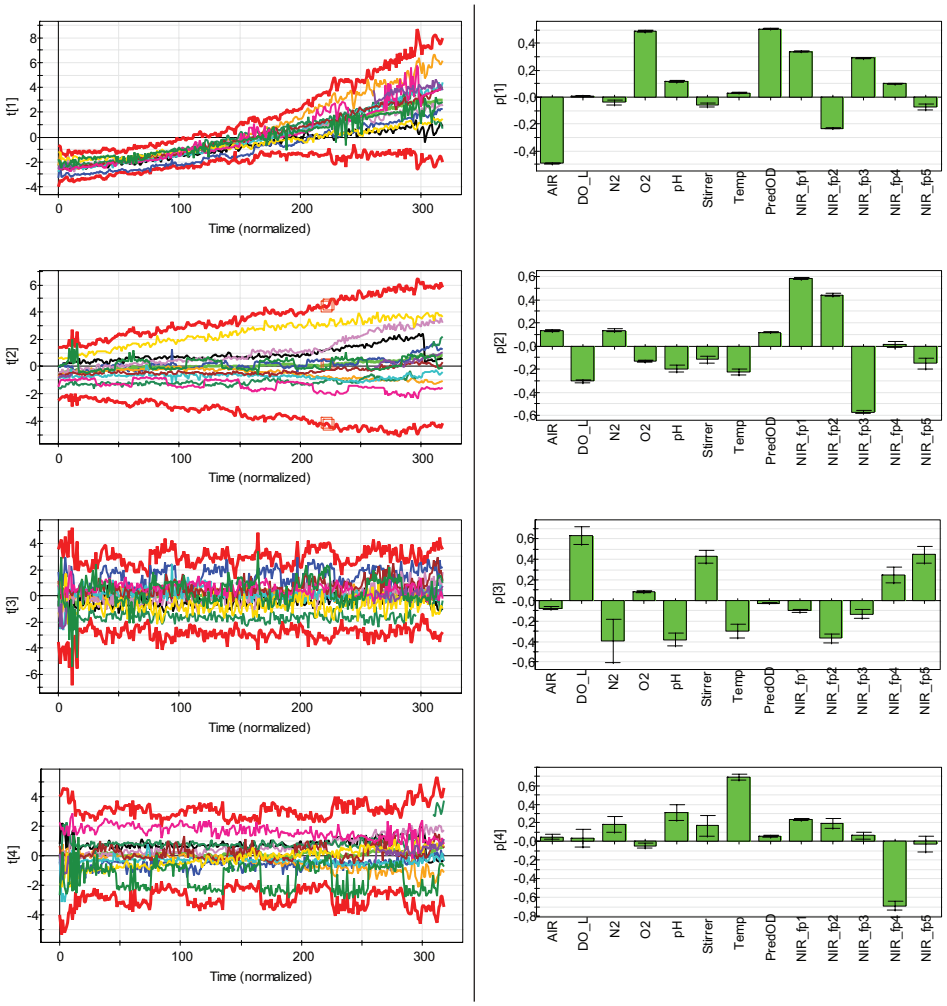


Figure 6. Scores (left) and Loadings (right) of Phase 2 of components 1 to 4 (top to bottom) of the dynamic evolution model. The outer red lines in the line graphs indicate a 95% confidence interval and the green line the average. The other lines each represent the evolution of the score of a cultivation batch over time. The loadings (right) are represented as contribution plot in which the contribution of each individual parameter to the variance is shown. Error bars indicate the standard deviation over time.

the cultures run at 35 °C(circles) and at the left and top part the cultures run at 37 °C. Also a separation in the rector used can be seen with the runs in reactor A (open symbols) being higher in the plot than the runs in Reactor B (closed symbols).

Because all batches were concluded to be of good quality, this plot – including the ellipse at

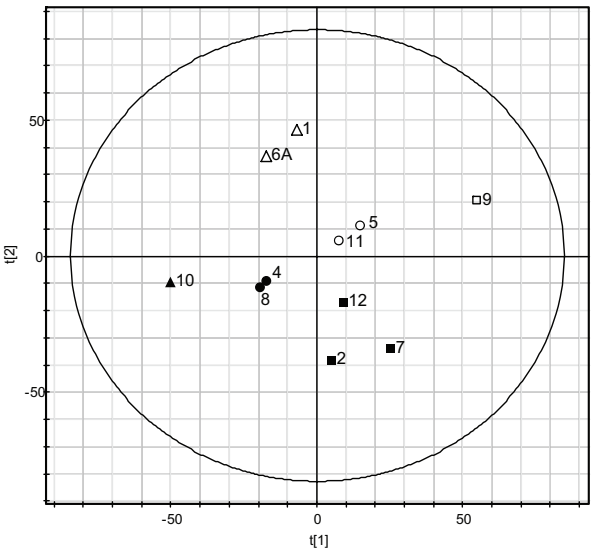


Figure 7. Overview of the multivariate batch monitoring model. Batches marked with triangles were run at 37 °C, with circles at 35 °C and with squares at 33 °C. Open symbols represent the batches run in Reactor A and closed symbols batches run in Reactor B. The numbers denote the batch numbers as denoted in Table 1.

the 95% confidence interval – can be viewed as a design space. However, the present dataset is insufficient to allow release decisions based on this model, because the present model does not fully describe the process design space. A batch with a combination of variables that is not tested in the DoE setup may fall outside the current design space, but still be of good quality or, on the other hand, may fall inside the design space and be of insufficient quality. As mentioned above, data from new batches must be added to the model to increase its accuracy, especially near the edges of the process design space.

Conclusions

This report demonstrates how the general principles of PAT and design space can be applied to undefined biopharmaceuticals. The complexity of a bacterial cultivation process first requires sound scientific investigation of the critical process and product attributes and the analyses needed to measure these attributes. Mandatory QC assays may not provide the exact scientific information necessary to appraise product quality. A “pass” or “fail” result is usually insufficient in this case. The quality of the bacterial suspension at the end of cultivation was appraised using a product quality score derived from DNA microarray analysis. This provides a quantitative measure of the performance of the run. The already available on line data were supplemented with NIR spectroscopy and samples were taken

frequently to determine nutrient and biomass concentrations. This whole set of data contains all relevant information for this process step.

This report shows how this complex data set can be integrated into a single process model that descibes the process design space for the cultivation of *Bordetella pertussis* bacteria for the production of a whole cell vaccine against whooping cough disease. The steps that were undertaken for the development of this model are described; from experimental design and analytical (PAT) methodology to chemometrics and mathematics. The resulting model can be used to monitor process evolution on-line and make predictions on the expected product quality at the end of cultivation.

Future application of this approach to all manufacturing process steps will allow the development of an overall manufacturing model incorporating the input and output of each process step, thus enabling flexible manufacturing at optimal settings. When this is validated, any process that has been executed within the process design space is assured to be of the specified quality. This means that this process model can make the release decision of the product during processing. Of course, to be allowed to replace mandatory QC testing, the model must be confirmed through extensive testing using actual process data. The model presented in this paper may be considered a starting point for this procedure. It should be run in parallel to the normal testing procedures and, when sufficient data is added to the model and regular tests have confirmed its reliability and comparability, it can eventually replace these tests. The resulting system is then capable of *…analyzing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality*, fulfilling the PAT definition.

Appendix 1: Interpretation of the model components

Phase 1

The first component represents a feature that increases throughout the batch evolution. It consists of a correlation between main variables AIR, stirring, PredictedOD and the first NIR fingerprint (NIR_fp1), which is mainly the offset that also explains OD. These all increase while N_2 is decreasing. This is a logical correlation since they all follow general trends caused by bacterial growth during batch cultivation. Increases in oxygen consumption cause the fraction AIR to increase over the fraction pure N_2, and combine with an increasing stirrer speed and of course increased optical density.
The second component consists mainly of a correlation between temperature and NIR_fp4. This is simply a confirmation that temperature is reflected in the NIR spectra.
The third component shows AIR, DO and O_2, together with NIR_fp4 decreasing, while N_2, pH, Temperature and NIR_fp3 increase. This might just be an effect caused by the different DO setpoints in combination with temperatures in each batch. Although dissolved oxygen (DO) is stable throughout each batch, between batches there are differences between the N_2/AIR ratios needed to maintain the setpoint. When the setpoint of DO is low, more nitrogen and less AIR is needed. This component also picks up an increase in pH that occurs mainly at the end of Phase 1. These phenomena maybe be caused by the fact that the oxygen transport constant between liquid and gas (k_La) is dependent on temperature. At lower temperatures k_La is lower, which means a higher oxygen concentration in the gas phase is needed to maintain the setpoint. Also the solubility of CO_2 (as H_2CO_3) is increased which results in a reduction in pH.
The fourth component shows that O_2 is negatively correlated with primarily NIR_fp2. Close examination of the scores plot shows that this component reflects a difference between the reactors – batches run in Reactor A have low score values and batches run in Reactor B have high values. The variables NIR_fp2 and NIR_fp3 reflect some reactor difference (having higher values for Reactor A) and since the variable O_2 has a small variation between reactors – varying between 0 and 0.0005 for Reactor A but between 0 and 0.0015 for Reactor B, this component picks up a small negative correlation between the variable O_2 and the two NIR fingerprint variables.

Phase 2

The first component, again, exhibits a rising trend for all batches. It consists mainly of a correlation between PredictedOD and increased NIR_fp1, while AIR decreases and O_2 increases. This indicates an increased oxygen fraction in the gas flow due to the mixing of pure oxygen while reducing the AIR to maintain a constant gas flow. This phenomenon is accompanied by a strong increase in the optical density and it is reflected in the NIR_fp1.

The second component shows a diverging pattern with some batches increasing while some decrease slightly. The correlation between the factors is relatively weak, except for the NIR fingerprints, where NIR_fp1 and NIR_fp2 strongly increase and NIR_fp3 decreases. For the other factors, AIR and N_2 and PredictedOD increase while all others (O_2, pH, Stirrer, Temperature) decrease. This complex interaction cannot be readily explained but may be the result of several effects due to both bacterial growth and the $k_L a$ phenomenon described above.

The third component shows a rather noisy pattern, which is reflected by the larger error bars in the loading plot. Predominantly, an increase in DO and Stirrer is accompanied by a decrease in N_2, pH and temperature. The strong effect of pH is hard to explain, but the others are parameters that have been correlated earlier. This seems to be another difficult to explain interaction but it may be due to the influence of several effects arising from the different set points for all experiments in the design.

The fourth component shows a slightly less noisy pattern than the third, but it is still largely unstructured. It mainly involves a correlation between pH and temperature. This can be explained by the aforementioned phenomenon that CO_2 (which is produced by the bacteria in the fermentor) dissolves better at low temperatures and, since it dissolves as a weak acid (H_2CO_3), reducing the pH .

References

1. ICH. 2005. Pharmaceutical Development (Q8). International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use http://www.ich.org/LOB/media/MEDIA1707.pdf.
2. FDA. 2004. Guidance for Industry, PAT - a framework for innovative pharmaceutical development, manufacturing and quality assurance. http://www.fda.gov/cder/guidance/6419fnl.pdf
3. Harms J, Wang X, Kim T, Yang X, Rathore AS. 2008. Defining process design space for biotech products: case study of Pichia pastoris fermentation. Biotechnol Prog 24(3):655-62.
4. FDA. 1987 (2008 version). Guideline on General Principles of Process Validation. http://www.fda.gov/CDER/GUIDANCE/pv.htm.
5. ICH. 2004. Comparability of biotechnological / biological products subject to changes in their manufacturing process (Q5E). http://www.ich.org/LOB/media/MEDIA1196.pdf.
6. Montgomery DC. 2001. Experimental Design for Product and Process Design and Development. Journal of the Royal Statistical Society 48(2):159-177.
7. Streefland M, van de Waterbeemd B, Kint J, van der Pol LA, Beuvery EC, Tramper J, Martens DE. 2008. Evaluation of a critical process parameter: Oxygen limitation during cultivation has a fully reversible effectt on gene expression of *Bordetella pertussis*. Biotechnol Bioeng 102(1):161-167.
8. Wold S, Kettaneh N, MacGregor J, Dunn K. 2008. Batch Processes: Modelling, Analysis and Control. In: Walczak B, TaulerFerré R, Brown S, (editors). Comprehensive Chemometrics. Elsevier.
9. Thalen M, van den IJssel J, Jiskoot W, Zomer B, Roholl P, de Gooijer C, Beuvery C, Tramper J. 1999. Rational medium design for Bordetella pertussis: basic metabolism. J.Biotechnol. 75(2-3):147-159.
10. Kasahara T, Miyazaki T, Nitta H, Ono A, Miyagishima T, Nagao T, Urushidani T.. 2006. Evaluation of methods for duration of preservation of RNA quality in rat liver used for transcriptome analysis. J. Toxicol Sci 31(5):509-19.
11. Mutter GL, Zahrieh D, Liu C, Neuberg D, Finkelstein D, Baker HE, Warrington JA. 2004. Comparison of frozen and RNALater solid tissue storage methods for use in RNA expression microarrays. BMC Genomics 5(1):88.
12. Ramalho AS, Beck S, Farinha CM, Clarke LA, Heda GD, Steiner B, Sanz J, Gallati S, Amaral MD, Harris A and others. 2004. Methods for RNA extraction, cDNA preparation and analysis of CFTR transcripts. J Cyst Fibros 3 Suppl 2:11-15.
13. Cummings CA, Bootsma HJ, Relman DA, Miller JF. 2006. Species- and strain-specific control of a complex, flexible regulon by Bordetella BvgAS. J.Bacteriol. 188(5):1775-1785.
14. Streefland M, van de Waterbeemd B, Happe H, van der Pol LA, Beuvery EC, Tramper J, Martens DE. 2007. PAT for vaccines: the first stage of PAT implementation for development of a well-defined whole-cell vaccine against whooping cough disease. Vaccine 25(16):2994-3000.
15. Irwin JO, Standfast AF. 1957. The intracerebral mouse-protection test for pertussis vaccines. J.Hyg.(Lond) 55(1):50-80.
16. Kendrick PL, Eldering G, Dixon MK, Misner J. 1947. Mouse Protection Tests in the Study of Pertussis Vaccine: A Comparative Series Using the Intracerebral Route for Challenge. Am J Public Health Nations Health 37(7):803-10. 10. Kasahara T, Miyazaki T, Nitta H, Ono A, Miyagishima T, Nagao T, Urushidani T. 2006. Evaluation of methods for duration of preservation of RNA quality in rat liver used for transcriptome analysis. J Toxicol Sci 31(5):509-19.
17. Xing D, Das RG, O'Neill T, Corbel M, Dellepiane N, Milstien J. 2001. Laboratory testing of whole cell pertussis vaccine: a WHO proficiency study using the Kendrick test. Vaccine 20(3-4):342-51.

Chapter 7

General discussion

Submitted in a revised form to Trends in Biotechnology

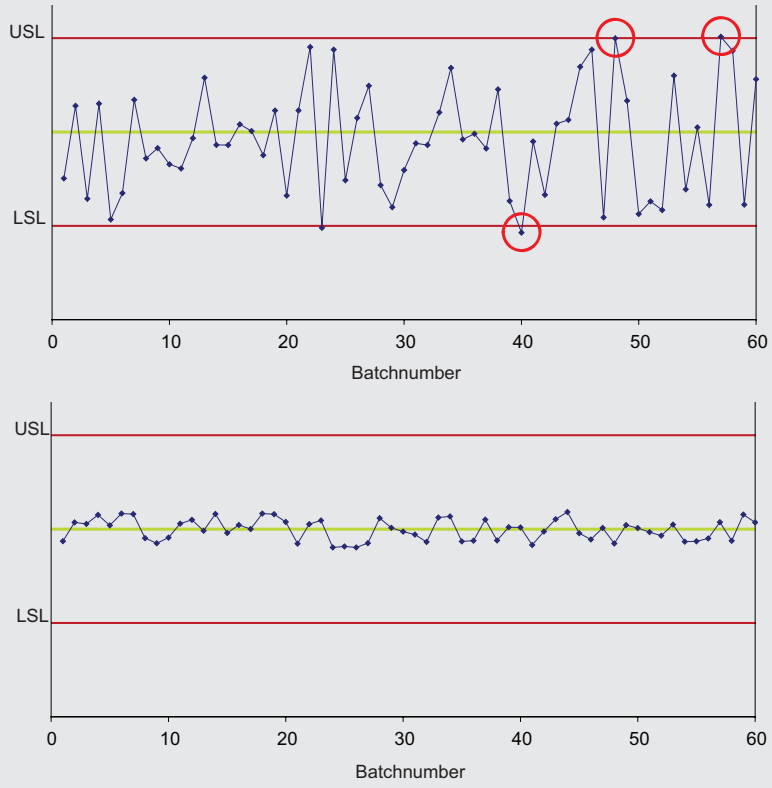
Introduction

This thesis describes the application of Process Analytical Technology (PAT) on the bioreactor process step for a whole cell vaccine against whooping cough disease. A whole cell vaccine is probably one of the most complex (bio)pharmaceutical products and therefore the most challenging for application of PAT. However, the work described in this thesis demonstrates that even for such an undefined product, PAT application is feasible and that PAT can thus in principle be applied to any (bio)pharmaceutical process. PAT application on any pharmaceutical process step can be divided in six steps, which are given below, exemplified with the work presented in this thesis:

1. **Determine the critical product attributes**. This means that a (preferably mechanistic) understanding of the function of the product in the human body is acquired. Also the specific parts or functional groups responsible for both therapeutic and adverse effects should be known (Chapter 2).
2. **Develop assays, analyses or sensors to quantitatively measure these critical product attributes.** Ideally this information should be available on line as a PAT tool (Chapter 2 and 4). For the relatively undefined whole cell vaccine against whooping cough disease, potency is routinely determined using a mouse challenge assay[1]. Such an assay is highly variable and gives only a qualitative result regarding product quality (pass or fail). This information is not suitable for the development of a PAT application. In this thesis an alternative method for the determination of vaccine potency is proposed based on DNA microarrays. In this way a quantitative measure for potency is acquired, which allows correlation between critical process attributes and critical product attributes.
3. **Acquire understanding of the interaction between the process and the the critical product attributes.** Identification of the critical product attributes (Chapter 2) and investigation of the critical process attributes (Chapter 3 and 4) can only be achieved with sufficient understanding of the underlying (cellular) mechanisms that define product quality. The process parameters that are critical for product formation and their tolerance limits need to be known (Chapter 3). This includes both physical (e.g. temperature, shear) and chemical (e.g. pH, substrate concentrations) data on the product environment.
4. **Incorporate process analytical technology (PAT).** A quantitative determination of all process attributes that are critical for product quality needs to be available on line, or at least fast enough so that it can be used for process control. This can either be a sensor that directly measures critical product attributes or a multivariate (spectroscopic) sensor that captures process fingerprints. For the *B. pertussis* process no sensors that directly measure the critical product attributes on line are available, meaning the PAT application described in this thesis was based on indirect measurements. The direct measurement of product attributes using microarrays was used at the end of the batch to quantify product quality. The on-line measured parameters like the oxygen, nitrogen and air flow, agitation

Box 1. Six Sigma and Shewhart's control charts

The buzz-word Six Sigma originates from Motorola in the 1980's. It is based on statistical process control principles that originated much earlier with the work of people like Deming, Shewart, Ishikawa and Taguchi. The term refers to the expression of average manufacturing results with respect to the process specification limits. One of the most easy ways to illustrate this, is by using Shewhart's control charts (see figure below).



Shewhart's control charts. The chart shows manufacturing outcomes per run. Runs are chronologically plotted from left to right. The green line indicates target manufacturing and the red lines the upper and lower specification levels (USL/LSL). The upper panel shows a process running at a 2-3 sigma level, comparable to the current (bio)pharmaceutical industry. This means that 70-95% of production runs fall inside specification limits. Such a process has a significant chance for a faulty production (red circles). The lower panel shows a process running at a 6 sigma level. This process is called a "world class" process and has only 3.4 faulty runs per 1 million correct ones.

These charts show manufacturing outcomes over time and can be used to investigate if a process is still centered around the target outcome and how much variance the process has. It also includes the upper and lower specification limits. Manufacturing results above the upper or below the lower specification limit have to be rejected or reworked.

The sigma-level of manufacturing is based on these charts. First a Shewart's control chart is plotted. Than the standard deviation (sigma) of the process can be calculated. The number of "sigma's" that fit between the average and the upper or lower specification limits indicates the "sigma level" of the process. Six sigma would mean only 2 faulty productions for every 1 billion correct ones. However, a stable process (Fig. 1) is allowed to have a systematic variation of 1.5 sigma over time, which means that the process is not always averaged on the center but moves 1.5 sigma up and down around the center during the manufacturing lifetime of a product. Therefore, "six sigma" is actually "four-point-five sigma", which means 3.4 faulty productions for every 1 million correct ones.

speed and NIR data were used for this process to fingerprint conditions that lead to good product quality. (Chapter 5 and 6). Finally, near infrared (NIR) spectroscopy was found to be a promising PAT tool for the monitoring of the cultivation of *Bordetella pertussis,* the causative agent of whooping cough disease (Chapter 5). NIR spectroscopy can be used on line to quantitatively predict critical process attributes using partial least squares (PLS) models or qualitatively by principal component analysis (PCA), which gives a process fingerprint.

5. **Determine the process design space.** The ranges of critical process attributes that allow the manufacturing of product with the specified quality should be known, including the interaction of these parameters. This results in the determination of the process design space[2, 3]. In this thesis the known critical process attributes (Temperature, pH, DO, density of the seed culture and inoculation density) were studied using a design of experiments approach (Chapter 6). This study was at the same time used to obtain process fingerprints that lead to proper product quality as described in point 4.
6. **Develop a model that describes the process design space**. The process data that constitute the process design space need to be available on line so that the process control system can match information from PAT sensors with the design space data to check if the process is running on target (Chapter 6). This control system needs to signal if the process starts to run off target. In a next step the process control system should automatically adjust the process so that it returns on the target trajectory. In this way product quality is ensured during processing, which is one of the PAT principles. For the pertussis vaccine process, the NIR data, the regular process data (i.e. pH, DO, temperature) and the microarray data were finally integrated in a statistical model that describes the process design space for the cultivation of Bordetella pertussis for a whole cell vaccine against whooping cough disease (Chapter 6). Although this model is only a starting point and needs to be complemented using data from actual manufacturing runs, it demonstrates the feasibility of PAT application on a complex product like a whole cell vaccine.

Thus by implementing these six steps for the fermentation step of the pertussis vaccine production process for the first time an initial PAT application has been developed for a complex biopharmaceutical product For other biopharmaceuticals PAT applications have not been developed to date. In general, the more complex the product, the harder it is to develop a PAT application for such a product. Modern biopharmaceuticals such as monoclonal antibodies are already much better defined than, for instance, a whole cell vaccine, but they remain complex structures created using complex processes. However, as long as the critical product attributes of any product are known, the interaction between process parameters and these parameters are mechanistically understood, and a process sensor is

available that captures the critical process attributes, PAT application should be feasible for any biopharmaceutical product.
In this chapter the current status of PAT for biopharmaceutical processes will be disussed together with the challenges for developing PAT for different biopharmaceutical products and production platforms. This will be done specifically for the bioreactor process step, because this is where the product is being formed and therefore this step can be considered the most critical process step, both for a whole cell vaccine as well as for a biopharmaceutical protein.

Process quality in the (bio)pharmaceutical industry

The design of processes in any manufacturing industry requires understanding of engineering principles combined with understanding of the mechanism of the product function and assembly. For most products this engineering perspective will result in robust processes that allow consistent production of high quality products. Sometimes these can even be called "world class" processes, fulfilling the six-sigma principle (see Box 1). This means only 3.4 faulty productions for every one million correct ones for a stable process.

The current situation in the pharmaceutical industry is shown on the left panel of Figure 2: a sigma level between 2 and 3. A sigma level of 3 means 5% faulty productions. Application of PAT is aimed at increasing the industry's sigma level. A sigma level between 4 and 5 is considered to be feasible for (bio)pharmaceutical products.

One of the most cited persons in the field of process quality thinking is William Edwards Deming[4]. He was one of the first to introduce "statistical quality control", which comprises a set of tools to analyze and improve manufacturing processes, such as the process control chart (Box 1). Deming started after the Second World War in Japan, where his ideas helped this country to become the world leader in the manufacturing of affordable, high quality consumer goods throughout the 1960's and 1970's. Key principle was to reduce costs and increase process efficiency continuously.

Process stability and capability are essential concepts for efficient manufacturing and process development. A process is considered stable when mean, variance and distribution of the variance of product specifications are constant over time. A process is considered capable when it is able to consistently produce output within its specification limits. Both principals are illustrated in Figure 1.

Currently the (bio)pharmaceutical industry functions at a level between 2 and 3 sigma[5], which means that between 70 % and 95 % of all manufacturing falls within specifications, but 5-30%

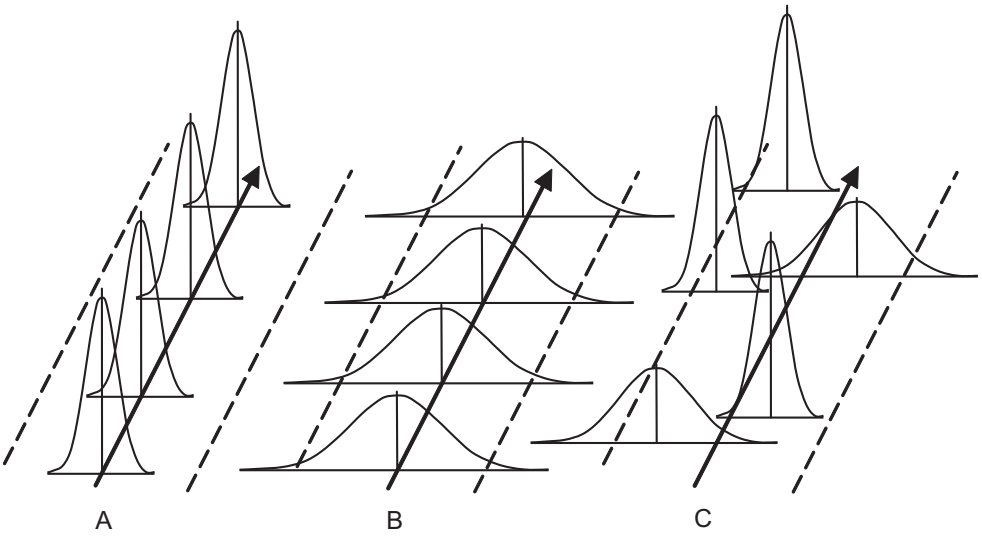


Figure 1. Process stability and capability. 3D representation of Shewart's control charts (Box 1). Each curve represents average processing and the corresponding variance of a certain time interval (i.e. a group of batches) of the total manufacturing lifetime. The arrow indicates the manufacturing life time of the product. **A**: stable and capable process. Average manufacturing remains centred over time and the variance indicates that product quality remains well within the specification limits (dotted lines) **B**: stable but uncapable process. The process remains centred over time, however the variance indicates a significant chance that manufacturing output does not meet the specification levels. **C**: unstable and incapable process. The process variance should allow consistent manufacturing within the specification levels, however, the average process output shifts over time towards the upper or lower specification level (dotted lines) causing manufacturing outside the specification levels and subsequent rejection

falls outside specification limits (Box 1). The quality of the output of (bio)pharmaceutical processes is ensured by extensive testing of the final product in order to prevent release of *out of spec* products. Because this testing takes place on finished products instead of during processing, rejection of out of spec products are a considerable cost component of (bio)pharmaceutical manufacturing.

This inefficiency is caused by a combination of a highly regulated industry with poorly monitored and controlled processes. Good manufacturing practice (GMP) regulation relies on documented quality assurance based on end product testing. Process performance is usually not included in the quality system and the quality control (QC) tests that are used are not necessarily representative for safety and efficacy of the product in humans. Thus product quality is inadequately assured during processing and, as stated earlier, needs to be determined on the finished product. In addition to this, a detailed description of the process is part of the registration file of the (bio)pharmaceutical product. This means that when a

product is approved for marketing, the validated process specifications in the registration file can not be changed. Usually when new products have just entered full scale manufacturing, little is known about the processing capability and stability. Process variance is estimated based on a limited number of batches during early manufacturing and fixed upon approval of the process. All future manufacturing is bench marked on these batches and excursions from the approved process need to be formally reported as deviation to the regulatory authorities. For the manufacturing of biopharmaceuticals, this "process is the product" thinking makes an approved process essentially fixed. As a consequence, process optimization or changes to the process in general need extensive validation and testing in order to assure that the product remains unchanged or "essentially similar". When this can not be assured, new clinical tests are usually required, which makes process optimization a costly matter.
Finally, the regulation of the pharmaceutical industry is enforced through timely inspections of production facilities and careful evaluation of all documented deviations, (re)validations or other test results that support process changes and disturbances. The enormous amounts of paperwork that are accumulated for this purpose have reached a point that inspection becomes more and more difficult and time consuming for regulators as well as manufacturers.

The PAT Initiative

In an attempt to overcome this, the American Food and Drug Administration (FDA) launched the Process Analytical Technology or PAT initiative in 2002[6]. With this initiative, the FDA stimulates the use of current scientific knowledge, modern process analytical tools and general processing quality principles for design, analysis and optimization of manufacturing processes. Or, as formulated by the FDA: *PAT is a system for designing, analyzing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality.* (FDA's PAT definition). In other words, a system that ensures process capability and stability. Also, manufacturers are now allowed to collect manufacturing data for research purposes only, without having to show that data upon inspection. This means that manufacturers can investigate their processes without having to fear that unexpected or unwanted outcomes of the investigations can be used against them during inspection.

In the recent years, the PAT principles have been adopted by the European EMEA and are integrated in recent documents from the International Conference of Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). The ICH Q8, Q9 and Q10 guidances[7] have set the standard for pharmaceutical process development, manufacturing and quality assurance in the 21st century.

When manufacturing becomes more and more reliable, consistent processing can be

demonstrated (less deviations, less rejects) and product quality becomes an inherent trade of the process (real time quality assurance), the need for inspection will decrease. This is of course a major benefit for regulators. For manufacturers the main benefits lie in cost reduction by more consistent processing (less rejects), processes that are more optimal (higher yields) and can be faster optimized and, just as for regulators, a lower regulatory burden. Of course, the purchase of a PAT infrastructure and the acquisition of the required level of process understanding will remain an upfront investment, which initially will require extra time and money and thus can be reason for manufacturers to stick to the quality-by-testing paradigm. However, regulations are changing already and the ability to demonstrate scientific process understanding and quality assurance during processing will therefore soon require a PAT system that monitors "quality-by-design" processes.

PAT for biopharmaceuticals

For chemical drugs the PAT initiative has already resulted in several approved PAT applications (i.e. for tablet inspection or granulate drying). For biological processes, however, the upstream cultivation or fermentation unit operations, PAT application is considered especially difficult. This is mainly because of the difficulties involved in on-line measurements of critical product quality attributes of complex biopharmaceutical products. Recently, the FDA has opened an invitation for biopharmaceutical companies to work with the agency to submit a pilot PAT submission for a biopharmaceutical product[8], in an attempt to address the industry's reluctance to adopt PAT principles.

Biopharmaceuticals or biologicals are generally considered a separate group of pharmaceuticals. Especially the "original" biopharmaceutical products like vaccines that consist of whole inactivated viruses or bacteria, or products purified from blood plasma, constitute a relatively undefined group of therapeutics as compared to chemically synthesized small molecules. This is reflected in the fact that the American FDA has separate offices for approval of new biopharmaceutical products (CBER) and small molecules (CDER). Notably, biopharmaceutical products like monoclonal antibodies and recombinant therapeutic proteins have been found to be defined enough so that they can be subject to the same approval regulation as small molecules. Since a few years, these biological products are governed by the Office of Biotechnological Products (OBP), which is part of CDER. This makes approval of these products easier and faster, which aids the current developments of the pharmaceutical industry, because most of the newly approved biopharmaceutical products are monoclonal antibodies or therapeutic proteins[9].

Looking at the approved biopharmaceutical products of the past 4 years in the US and Europe (Table 1[10-13]) reveals that three production platforms are currently predominant: the bacterium *Escherichia coli*, the yeasts *Saccharomyces cerevisiae* and *Pichia pastoris* and the mammalian Chinese Hamster Ovary (CHO) cell line. Mammalian cells are emerging as the new workhorse of biopharmaceutical manufacturing, as is illustrated by the fact that in 2007, 12 out of the 15 approved biopharmaceutical products were manufactured using mammalian or insect cells[9].

Table 1. The number of approved biopharmaceutical products in the US and EU in 2004-2007 and their respective production platforms (From Walsh (2005-2008[10-13])

Year	Total no. of products approved	*E. coli*	*S. Cerevisiae / P. pastoris*	CHO	Other mammalian or insect cells	Other production platforms
2004	12	2	1	4	3	2
2005	10	4	2	4		
2006	13	4	2	3	2	2
2007	15	2	1	8	4	
Total	50	12	6	19	9	4

The manufacturing of biopharmaceuticals generally starts with a series of cultivations in order to accumulate enough biomass to inoculate a final production bioreactor. The bioreactor process step is then followed by a series of purification steps in order to obtain the purified product in the desired formulation. These steps do not add to the intrinsic quality of the product as it is formed during the cultivation step, but they are aimed at removing unwanted compounds in order to prevent side-effects and increase product safety. In general, the higher the necessary purification effort, the lower the recovery of the biopharmaceutical product. As stated above, the bioreactor process step can be considered the most critical process step in biopharmaceutical manufacturing and therefore the most important process step for PAT application.

Available PAT tools for bioreactor monitoring

Implementation of PAT means applying the correct sensor(s) to measure the critical product and process attributes of the bioreactor process. Depending on the critical attributes that need to be monitored during processing, many PAT tools can be applied either on line or off line. An excellent overview of the available tools is given by Kansakoski and coworkers[14]. A summary of currently investigated PAT tools for bioreactor analysis is given in Table 2.

Table 2 Available PAT tools and their area of application. Information adapted from[14-16]

PAT tool	Measured attribute	on/at/off line*
Optical Spectroscopic tools		
Near Infrared Spectroscopy	Many biologically important bonds (aliphatic C-H, aromatic or alkene C-H, amine N-H and O-H), physical properties (particle size, morphology, optical density)	on line
Infrared and RAMAN spectroscopy	Detailed chemical information of compounds	on/at line
Photoacoustic spectroscopy	off gas analysis	on line
2D fluorescence spectroscopy	Many biomolecules and co-factors with fluorescent properties	at / off line
UV spectroscopy	Measures absorption of biomolecules in the UV spectral range	on / off line
Real time imaging	On line microscopic analysis of the cultured organism	on line
Other spectrometric tools		
Mass spectrometry	Analysis of metabolome or proteome of the cultivated organism, to get information on the metabolic and physiologic state, expression of recombinant proteins, growth characteristics and harvest point determination	off line
Dielectric spectrometry	Analysis of membrane potential to assess the viability of the culture	on line
Biomolecular tools		
Green Fluorescent Protein	Can be used as a marker for product expression and easily measured	on line
Biosensors / Biochips	Sensors that are activated upon binding of a molecule to a receptor. Highly sensitive for specific compounds	at / off line
Transcriptome or proteome measurements (microarray)	Analysis of the transcriptome of the cultivated organism, which gives information on the metabolic and physiologic state, expression of recombinant genes and determination of optimal harvest point	off line
Other tools		
Flow cytometry	Analysis of cell morphology and expression of surface proteins and isolating sub-populations of cells	at / off line
On line HPLC	Analysis of a cell free sample stream, giving information about nutrient and metabolite concentrations during cultivation.	On line
Softsensors	Mathematical algorithms that perform calculations on one or more on line available parameters to calculate or predict a parameter that is not on line measurable	on line
Dynamic modelling and prediction	Statistical process models that incorporate a wide range of process data from PAT tools to describe the process and make predictions on product formation and quality	on line

* On line: real time measurements of the process; at line: off line measurements that are fast enough to be used to control the process; off line: off line measurements that can not be used to control the bioreactor process

The PAT tools described in Table 2 can be used for different types of measurements in bioreactor processes. Some allow direct on line measurements of critical attributes or directly linked parameters, like on line HPLC[17]. Others give on line multivariate data that needs chemometric processing before interpretation, such as spectroscopic tools[16] like as near infrared spectroscopy. Off line tools such as DNA microarrays can not directly be used for process control, but provide detailed information on the cellular biology of the production platform organism. All gathered knowledge and data finally need to be captured in a statistical relationship that describes the process and the evolution of critical product attributes during the process.

PAT tools for direct or indirect measurement of critical attributes

Direct measurements of critical attributes are preferable, when they are available. For most biological products, critical product attributes are measured using labor- and time intensive QC tests, such as ELISA, blotting techniques or animal tests. The information from these tests can rarely be used on line or at line, although the trend is towards faster, high throughput assays[18]. Recent developments in the biosensor and biochips fields[19, 20] show promising developments that would make on line measurements of more critical product attributes of complex biopharmaceuticals feasible in the near future. As for now, these sensors are usually too complex or unreliable to be used in controlling biopharmaceutical manufacturing.

In many cases derived, indirect parameters of critical attributes can be measured more easily than the critical attribute itself. For instance, the direct measurement of the capacitance of the cell membrane using dielectric spectrometry can be used as a surrogate for mammalian cell culture viability[21]. A direct measurement of viability involves sampling, staining and microscopic observation, which can not be made available on line. A number of robust sensors are available at the moment (i.e. Aber®'s capacitance probe). Also, computer algorithms (so called soft sensors) are available that predict a non-measurable parameter based on its interaction with parameters that can be measured on line. For instance, bacterial cell density can be predicted based on the culture's oxygen consumption rate [22,23]. The trend to bring measurements on line that were previously only available off- or at line allows a new level of feedback control based on these new on line process data. This will enhance the controllability of bioreactor processes.

PAT tools for multivariate analysis of critical process attributes

Multivariate measurements, like near infrared spectroscopy, are powerful in the sense that they can capture many variables in a single measurement. The difficulty is that it is usually hard to directly link critical attributes to one or more of these variables. Correlation between multivariate measurements and critical process and product attributes requires chemometric tools such as principal component analysis (PCA) or partial least squares (PLS) calibration models (see Chapters 5 and 6) . These techniques search patterns in the multivariate dataset

that correlate with the critical attributes of the process or the product. After validation, these correlations can be used as an indirect measurement of the critical attributes on line. This approach can be used for measurement of critical attributes for which no direct on line measurement is available. A number of robust probes for multivariate bioreactor monitoring are commercially available (i.g. NIR and Raman probes from several suppliers, or the fluorescence BioView® probe from Delta, Denmark). These probes can be readily implemented for bioreactor monitoring without technical adjustments to the bioreactor vessel. In that respect they form a good starting point for PAT application, because data collection can be started immediately, without disturbing the existing process. The collected data can then be analyzed (off line) for correlations with critical attributes and subsequently the PCA analysis or PLS calibration models can be run on line during manufacturing. However, these chemometric analytical tools require the aid of trained chemometricians, especially for the modelling of the more difficult non-linear phenomena, which might be a hurdle that prevents especially the smaller companies from implementing multivariate tools.

Transcriptome and proteome analysis

Transcriptome or proteome analysis using DNA or antibody microarrays gives complex data. Especially for eukaryote organisms or organisms that are not as thoroughly investigated as, for instance, *E. coli* or *S. cerevisiae*, interpretation of expression data can be difficult. Key is to identify the genes or proteins that are correlated with the critical attributes of the product and to use the available scientific knowledge of these specific genes and proteins regarding their role and function in cell physiology (see Chapter 2). In this way, an understanding is acquired about the interaction between cellular physiology and the product during manufacturing. For instance, how cultivation parameters can influence glycosylation profiles of complex biopharmaceuticals such as EPO (i.e. the genes and enzymes that are involved and how they are regulated). This understanding allows a more scientific approach towards process optimization and future process development. Another way the measurements of the transcriptome of the production organism can be used is as a fingerprint of the physiological (and metabolic) state of the organism. These fingerprints can be used to compare overall expression profiles of samples taken during manufacturing in order to investigate processing consistency. Of course, these techniques only show part of the complex picture of interactions between the several "omics" fields (i.e. transcriptomics, proteomics, metabolomics or fluxomics). Regulation of cellular processes can take place in may ways, some of which, like siRNA, we are only beginning to understand. Omics techniques will generally be less difficult to apply on simpler (prokaryote) organism than on, for instance, mammalian cells. The problem of data complexity can be partly overcome when only an isolated pathway or group of genes is important for a certain critical product attribute. Much like the product marker genes described in Chapter 2, only a limited number

of data points are investigated in detail, while the rest of the data can be used only to screen for deviating patterns. Of course this approach requires a priori knowledge of which genes are important for critical product attributes.

Descriptive process modeling

As`stated, key for PAT implementation is to establish a (statistical) relationship between processing conditions and the quality of the product. In order to be able to do this, the information from the bioreactor control system (i.e. pH, temperature, gas flows, etc) and from the additional PAT tools needs to be combined and correlated to information on product quality and scientific knowledge[24] (see Chapter 6). Such a model can then be used on line to monitor the process and check whether product of the specified quality is produced. In a second stage, these models should be able to control the process upon deviations or disturbances. The complete integration of product- and process understanding, process data and product quality with the aim to assure final product quality is what defines a PAT system.

Recently, software has come available that allows both the integration of all available data and the development and execution of process control models (Chapter 6 of this thesis). With this, the technical realization of full PAT application on a biopharmaceutical cultivation process has become feasible.

Challenges for PAT application in bioreactor processes

A bioprocess is considered more variable than a chemical process (e.g. chemical synthesis) or a physical process (e.g. drying). This assumption is based on the current common practice and the inherent theoretical complexity of a biological process. This is especially the case for the older biopharmaceutical products for which micro organisms and mammalian cells are cultured using complex cultivation media. Bovine blood serum, yeast extract, bovine heart extract or casein hydrolysates are examples of commonly used complex medium components. These components have a highly variable composition of a range of free amino acids, di- and tri-peptides and even more complex structures and co-factors. Batch to batch variation of a cultivation medium that contains these undefined components can be a major cause for bioprocess variation[25]. Currently, most of the production media are animal component free, protein free or even chemically defined, which makes the medium composition easily to monitor and control, a crucial prerequisite for PAT application

Another source of variance of a bioreactor process is the condition of the cells that are used to inoculate the final bioreactor. These cells usually originate from a frozen cell bank which is stored at -70 °C or -140 °C. This Working Cell Bank (WCB, for mammalian cells) or Working Seed Lot (WSL, for micro-organisms), is made periodically from a master seed or cell bank. The handling procedures to make a WSL or WCB from the master seed lot or cell

bank can already influence process variation[26,27].
Going from the WSL or WCB to the actual manufacturing scale bioreactor is another critical stage in terms of bioprocess variation. This usually involves a series of cultivations going from shake flasks or cultivation plates to small-scale bioreactors before the manufacturing scale bioreactor is inoculated. Ideally cells are maintained in logarithmic growth throughout this series, in order to maintain short lag times. The lag time is the time the cell needs to adjust to a new environment after inoculation[28]. Therefore, cells show shorter lag times when the conditions in the new flask or bioreactor resemble the conditions from the previous flask or bioreactor. In practice this means that the cells do not have to adjust their active metabolic routes when transferred into a larger vessel. When the new vessel is also at the same temperature and pH as the previous vessel, adaptation of the micro-organisms and with that the lag times can be reduced to a minimum[29,30]. The control of these factors can lead to reproducible bioreactor cultivations as has been demonstrated in Chapter 2.

Also, the strains or cell lines selected for production need to be stably transfected and remain genetically stable during the entire manufacturing chain from seed to the harvest of the large scale manufacturing bioreactor (ICH Q5D[7]). Furthermore, the cells need to be able to manufacture only product in the correct conformation, since it is very difficult to separate the correct product from its similar misfolded or differentially glycosylated counterparts during purification.
A good scientific understanding of the cell's physiology and the means to investigate the evolution of cellular physiology during manufacturing are helpful, if not crucial, for understanding the relationship between critical process and critical product attributes.

Biopharmaceutical bioreactor platforms
As mentioned above, the majority of biopharmaceutical processes involve the cultivation of either *E. coli, S. cerevisiae* or a CHO cell line. Although general bioprocess engineering principles may apply on all of these organisms, they have very different and specific demands and characteristics that need to be considered in order to develop a process compliant with the PAT principles. The choice of the expression system is usually based on the specific characteristics of the therapeutic protein that is being expressed[31]. Proteins that require (eukaryote) post-translational modifications such as acetylation, amidation, carboxylation, hydroxylation, disulfide bond formation, glycosylation, phosphorylation, proteolytic processing or sulphation[32] in order to attain their full biological and therapeutic function can, for instance, not be effectively produced using the prokaryote *E. coli*, but require mammalian cells. *E. coli* is highly suitable for the expression of small proteins or peptides, however[9, 33]. The yeast *S. cerevisiae* can serve as an intermediate, in the sense that it is able to perform eukaryote post translational modifications. However, the obtained modification pattern does

not resemble the human pattern, which may give problems in proteins that are strongly post translationally modified[34].

Bacteria

Bacteria have the obvious benefits of fast growth on simple and cheap media and can over express large amounts of recombinant proteins. Furthermore, they can be easily transfected to yield genetically stable production strains. This makes them commercially very attractive.
The cultivation of bacteria for biopharmaceutical production generally involves either a batch or a fed-batch strategy. A batch cultivation is technologically the most simple cultivation strategy, which means it also has, in principle, the least number of critical process attributes. For batch cultivation the starting material, i.e. the seed and the medium, is one of the two key parameters to be controlled (Chapter 2), because during cultivation only basic parameters such as pH, dissolved oxygen and temperature can be controlled, usually at a fixed set point. This means that these parameters are only critical for product quality in case of sensor failures or process disturbances (Chapter 3). The other key parameter to be controlled is the determination of the harvest point. It can be critical to harvest before nutrients become limiting and physiological changes in the cell compromise product quality (Chapter 4). Or, more commonly, cell lysis starts after the nutrients are depleted, causing massive release of DNA and cellular proteins such as proteases, which can hamper product purification or cause problems with product degradation.

In a fed-batch process, a nutrient containing feed is pumped into the bioreactor, allowing much higher biomass and product concentrations compared to batch cultivation. It can be operated either by supplying ample nutrients in order to ensure the maximum growth rate, or by supplying a limiting amount of nutrients and in this way control the growth rate with the feed rate[35]. The latter means that the bacteria are grown under continuous nutrient limiting conditions, which may have unwanted side effects on cell viability and subsequently on product formation[36, 37]. The critical attributes to consider for PAT application in a fed-batch process are not only seed and medium quality and controlling the physiological parameters pH, DO and temperature, but also the feed rate and feed composition. A fed batch cultivation usually ends when the desired optical density is reached, or, more practically, when the bioreactor vessel is full. The end of the process can therefore be easily monitored.

The high biomass densities that bacterial cultivation can reach, especially using fed batch, put high demands on the oxygen supply and mixing of the bioreactor vessel. This means that, especially at larger volumes, a bioreactor may suffer from gradients radiating away from where oxygen enters the bioreactor (usually near the stirrer). Gradients in pH and temperature are common in larger vessels as well. As a result of these gradients, the content

of the bioreactor can be inhomogeneous, introducing variance in bacterial physiology and the overall gene expression profile[38-40]. Understanding how gradients affect the product is a key aspect of reducing variance within a batch or fed-batch cultivation.

Because *Escherichia coli* is probably the most studied organism in molecular biology, it is easier to acquire a scientific understanding of the interaction between the process, the cell's physiology and subsequently product formation and quality. Much research has been done already on the optimization of recombinant protein expression in *E. coli* [41-43]. Because *E. coli* can only excrete proteins to a limited extent and mostly not beyond the periplasmatic space[44], high yield production of recombinant protein in *E. coli* usually involves the formation of inclusion bodies, which are intracellular aggregates of the over expressed protein. These aggregates need to be refolded after primary recovery in order to obtain the protein in its biologically active form[45]. This refolding is a very critical parameter for the conformational structure and thus the efficacy of the product, which is difficult to control. Thus, in terms of PAT the proper control of the refolding of proteins is a challenge for these processes.

When proteins are accumulated in inclusion bodies, induced expression is more favorable than constitutive expression, because the massive expression of the recombinant protein and the high concentrations of aggregated proteins inside the cell can interfere with normal cellular physiology[46]. Induction can be achieved either by an increase in culture temperature (heat shock) or by the addition of a specific chemical, that triggers the activation of the promotor of the recombinant gene[47]. Controlling the induction of product formation adds another critical parameter to the process, however, because the moment of induction and the growth environment can influence product expression[48]. This means that PAT tools need to monitor each phase (growth and production) and determine the optimal point for the induction of product formation[49,50].

This means that the challenges for PAT applications on bacteria, such as the control of process inputs, control of feed rate and composition, knowledge on process gradients and control of protein refolding, can technologically be met. The final challenge is to combine this technology with the appropriate scientific knowledge to build a PAT application that is capable of monitoring all critical process attributes through these PAT sensors and use this data for real time quality assurance of the product.

Yeast

S. cerevisiae and other yeasts, like *Pichia pastoris*, are also fast growing micro oganisms that can express large amounts of recombinant protein. Yields are typically somewhat less than an *E. coli* based process, but still commercially very attractive. Most of the PAT challenges for the cultivation of *E. coli* are applicable to yeast cultivation as well[51]: Yeasts can be cultivated

using either a batch or fed batch process and larger scale processes are prone to inhomogenities caused by gradients inside the reactor vessel. The more complex metabolism of *S. cerevisiae* and other yeasts makes the control of the feed for a fed batch process more difficult. Yeasts are generally able to utilize several carbon sources both aerobically and anaerobically[52, 53] This complex metabolism can cause *S. cerevisiae* to produce ethanol (usually an anaerobic metabolite, which can be toxic at millimolar concentrations) under aerobic conditions when excess carbon sources are available. Control of the feed rate and composition are therefore more critical and complicated as compared to a bacterial process.

Yeast can be easily and stably transfected for expression of recombinant proteins. Both constitutive and induced expression can be used, depending on the specific product characteristics. Recombinant human proteins, such as insulin, glucagon or granulocyte macrophage colony stimulation factor (GM-CSF) expressed in *S. cerevisiae* are among the first biopharmaceutical products on the market[34]. Because of its capability to make some post-translational modifications, it is more suitable to produce complex biopharmaceuticals than bacteria. However, the pattern of, for instance, N-linked glycosylation is markedly different from human glycosylation patterns[54] with glycosylation in yeast usually showing a high mannose content. Therapeutic proteins with a non-human glycosylation pattern, usually have a shorter plasma half-life time than proteins with a human glycosylation pattern. More importantly, they can become a target for the immune system with all sorts of unwanted side-effects as a result. To overcome this problem, the glycosylation pathways in yeast can be "humanized". For therapeutic proteins with relatively simple glycosylation patterns this can already be achieved, which means that yeast will become a more important production platform in the future. However, product yield and capping of N-glycans with sialic acid are challenges that still need to be resolved. An overview of the work in this field is given by Wildt and Gerngross[55]. For PAT application, the understanding of how processing conditions influence glycosylation profiles is the real challenge. PAT tools can be applied to monitor and control the processing conditions that will result in correctly glycosylated, or otherwise post translationally modified products.

Mammalian cells

Mammalian cells, such as the Chinese Hamster Ovary (CHO) cell, grow very slowly compared to bacteria or yeast. From a commercial perspective, this means that the platform is only interesting for products with a high added value that can not be produced using bacteria or yeast.

The strategy for cultivation of mammalian cells first depends on the cell's need for a solid surface for anchorage and growth. In order to be able to grow anchorage dependent in suspension, a solid surface is suspended in a bioreactor; the so called microcarriers[56].

Although this technology allows large scale cultivation of animal cells, sub-cultivation requires detaching of the cells and subsequent reattachment to the new surface. This is highly variable and labour intensive and is more critical than the sub-cultivation of suspension cells in terms of process consistency. Most new biopharmaceutical products are therefore produced using anchorage independent or suspension cells, such as the CHO cells. Therefore we will focus on these cells.

Because of the relatively low density of CHO cell cultures, large bioreactors with a working volume between 1 m^3 and 20 m^3 are used in order to have a high enough yield of the therapeutic protein[57]. The costs of large scale pharmaceutical grade bioreactors and the economic risk of a failed batch at even larger scale, prevent the use of even larger bioreactors. Manufacturers usually end to building a series of smaller bioreactors to reach the desired cultivation volume.

The complexity of mammalian cells results in many critical cultivation process parameters, starting with the medium composition. Even chemically defined media are relatively complex and only available for a limited number of cell lines. Furthermore, their genetic complexity makes it difficult to generate stable cell lines, as mentioned earlier. Also the selection of high producing clones[58, 59] and the requirement to demonstrate that the cell line originates from a single clone (ICH Q5D guidance[7]), are challenges for mammalian cells and even this does not guarantee a stable and reproducible production platform[60].

Fed batch or perfusion cultivation strategies are also applied in order to increase cell density. Wheras a fed-batch only has a feed rate, a perfusion culture has next to a continuous input of fresh medium also an outgoing cell free perfusion flow. In addition, often a bleed flow of cells and medium is required to maintain the viability of the culture above a certain level. The perfusion and bleed flow exactly balance the incoming feed rate, which means that the cultivation volume is constant. Furthermore, a steady state is reached after a certain time. Also the composition of the feed is different. In fed batch this is usually a concentrated solution of essential nutrients, while in perfusion it is usually the same as the cultivation medium. Both strategies introduce difficult to control critical process parameters. For example, the lack of understanding of cellular physiology and metabolism, at least when compared to the state of the art for *E. coli or S. cerevisiae*, makes it difficult to develop a feed with the optimal composition for fed-batch as well as for perfusion processes.
The retention of cells in a perfusion process, especially at larger scales or higher cell densities, is difficult to control resulting in variable bleed rates and as a consequence variable cell densities and viabilities. Furthermore, many cell retention devices are placed outside the controlled bioreactor environment, exposing the cells to gradients, which again are difficult to control and the effect of which on cell physiology is unknown.

Especially for perfusion processes, production times increase considerably and, with that, the number of cell divisions during manufacturing. Longer manufacturing times pose challenges for the stability of sensors for process monitoring and the sterility of the system. In addition, mammalian cell lines are in general less stable than yeasts and bacteria, which can have negative consequences for the stability and capability of the process. For example, the metabolism of a cell may change after a certain number of cell divisions, which will have consequences for the feed composition, for instance. However, stability of the cell line in terms of critical product attributes, such as glycosylation patterns of the product, needs to be demonstrated for approval of a new product (ICH Q5D[7]). Because the knowledge on what influences the stability of animal cell lines is limited, this usually means a trial and error approach for finding the optimal high producing clone, rather than a targeted search.

An understanding of the influence of process conditions and nutrient or metabolite concentrations on critical product attributes is necessary[61] for the application of PAT tools. This understanding is beginning to arise, especially for CHO cells, where genetic engineering and process optimisations allow the control of product glycosylation patterns[62]. This means that PAT application based on actual understanding of critical parameters of a mammalian cell process may be feasible in the near future.

An alternative for the engineering and selection of a stable, high-producing clone is transient expression. Transient expression can be achieved by both viral and non-viral gene delivery systems[63]. A recently approved transient expression system for biopharmaceutical products is baculovirus transfection of insect cells[13] (Cervarix®). This system is also becoming available for mammalian cells[64]. The main benefit of transient expression is the speed at which the transfection vectors can be generated and entered into the host cell for gene expression. Especially for, for instance, a fast track to clinical testing this method is widely used. However, stable transfection is preferable, both from a regulatory as from an engineering perspective, because the infection or transfection methods that are applied are very complex and hard to controll to ensure final product quality. Despite these challenges, mammalian cells are the only single-cell production platform that is able to express complex proteins with a human (like) pattern of glycosylation an other post translational modifications at acceptably high levels, which explains their increasing popularity as the workhorse of newly approved biopharmaceuticals[9, 13].

Conclusions

This thesis demonstrates how PAT can be applied on a bacterial cultivation process and how the critical process and product attributes of a complex biopharmaceutical can be assessed. For other biopharmaceutical production processes, the same principles can be applied, but each production platform and each product has its own specific critical attributes. Understanding of the interaction between critical process and product attributes is the starting point for any PAT application. For the development of a PAT system, the selection of the appropriate PAT tools, which are the tools that are able to monitor the critical attributes on line, is depending on the specific characteristics of each process and product.

One of the difficulties of bioreactor processes, especially for batch processes such as the cultivation of *Bordetella pertussis* described in this thesis, is that only a few parameters can actually be controlled during processing. For a batch cultivation these are usually pH, DO, temperature and stirring speed. Because of the lack of control during processing, the other critical parameters, such as seed quality, medium quality and composition or sterility, need to have been controlled at the start (inoculation) of the process. The only remaining critical parameter is the harvest point, which can be determined based on the monitoring of the evolution of critical process parameters (such as nutrient concentrations, see Chapter 4).

In general, each manufacturing platform has its own specific challenges. While bacteria can be stably transfected and easily grown to high densities at low costs, the limited range of products that are suitable for bacterial expression limits their use. Yeasts have largely the same benefits in terms of yield and ease of transfection as bacteria and they can express a much broader product range, especially when post translational modification pathways are “humanized”. However, medium and feed composition are more complex than for bacteria, because of the more complex metabolism, which introduces new critical parameters. Mammalian cells are the only suitable platform for high yield production of complex, highly glycosylated therapeutic proteins or antibodies. Their complexity, both on the genetic as on the metabolic level, introduces complex sources of variance and thus more challenges for PAT application.

When there are no means to compensate for variation in process input during processing (i.e. the process is fixed), there are two possible strategies. Either the variation in process inputs is reflected in process outputs, or, the process input variation is controlled to such an extent that process output becomes consistent. This second situation is common practice in current biopharmaceutical manufacturing and is described in Chapter 2. Application of PAT on such a process means monitoring of critical attributes during the entire process, from the moment the medium enters the bioreactor (before inoculation) until the harvest of the product. In this

way, process evolution can be captured and compared to the desired evolution trajectory or design space. This thesis describes the application of this approach on the cultivation of *Bordetella pertussis* bacteria for a whole cell vaccine against whooping cough disease.
This approach can be viewed as a first level of PAT application, where product quality is assured by strict compliance to operating procedures and real time confirmation by the process control system if everything is running as expected. The next level of PAT application involves a process that is capable of controlling all critical attributes. In other words, a variable process that adjusts itself in order to cope with variance in process input. Examples for this can be found in other industries. For instance, an oil refinery can cope with the high variation in the composition of crude oil, and still produce petrol and even more defined products with consistent composition. This is possible because the process adapts based on PAT measurements of the input materials.

For a bioreactor batch cultivation process this is much more difficult to imagine. Maybe it can look like a complex fed batch setup in which the medium components are all added depending on their specific consumption rates. Or even like a complex array of inducible genes that can be controlled by adding specific inducers depending on the evolution of critical process attributes. In either way, such an application depends on a sound scientific understanding of the mechanisms inside the cell that result in product formation and their interaction with processing conditions. It is this understanding that is the true enabler of PAT application and the extent of subsequent implementation is largely dependent on technological aspects like the availability of the appropriate sensors to measure a critical attribute or the process controller that can adjust the process accordingly. Already when the people that operate the process understand how processing conditions influence product formation and quality, will processing consistency increase.

References

1. Kendrick, P.L., *et al*. (1947) Mouse Protection Tests in the Study of Pertussis Vaccine: A Comparative Series Using the Intracerebral Route for Challenge. *Am J Public Health Nations Health* 37, 803-810
2. ICH (2005) Pharmaceutical Development (Q8). *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use* http://www.ich.org/LOB/media/MEDIA1707.pdf
3. Harms, J., *et al*. (2008) Defining process design space for biotech products: case study of Pichia pastoris fermentation. *Biotechnology progress* 24, 655-662
4. Deming, W.E. (1982) *Quality, productivity, and competitive position*. Massachusetts Institute of Technology, Center for Advanced Engineering Study
5. IBM (2005) The metamorphosis of manufacturing. From art to science. http://www-935.ibm.com/services/us/imc/pdf/ge510-4034-metamorphosis-of-manufacturing.pdf
6. FDA (2004) Guidance for Industry, PAT - a framework for innovative pharmaceutical development, manufacturing and quality assurance. http://www.fda.gov/cder/guidance/6419fnl.pdf
7. ICH (2008) Quality Guidances. http://www.ich.org/cache/compo/363-272-1.html
8. FDA (2008) Submission of Quality Information for Biotechnology Products in the Office of Biotechnology Products; Notice of Pilot Program. *Federal Register* 73, 37972-37974
9. Walsh, G. (2006) Biopharmaceutical benchmarks 2006. *Nature biotechnology* 24, 769-776
10. Walsh, G. (2005) Biopharmaceuticals: Approvals and Approval Trends in 2004. *Biopharm Int* 18, 58-65
11. Walsh, G. (2006) Biopharmaceuticals: Approvals and Approval Trends in 2005. *Biopharm Int* 19, 58-68
12. Walsh, G. (2007) Biopharmaceuticals: Approvals and Approval Trends in 2006. *Biopharm Int* 20, 40-48
13. Walsh, G. (2008) Biopharmaceuticals: Approvals and Approval Trends in 2007. *Biopharm Int* 21, 30-36
14. Känsäkoski, M., *et al*. (2006) Process analytical technology (PAT) needs and applications in the bioprocess industry http://www.vtt.fi/inf/pdf/workingpapers/2006/W60.pdf
15. Sitton, G., and Srienc, F. (2008) Mammalian cell culture scale-up and fed-batch control using automated flow cytometry. *J Biotechnol* 135, 174-180
16. Ulber, R., *et al*. (2003) Optical sensor systems for bioprocess monitoring. *Analytical and bioanalytical chemistry* 376, 342-348
17. Kochanowski, N., *et al*. (2006) Intracellular nucleotide and nucleotide sugar contents of cultured CHO cells determined by a fast, sensitive, and high-resolution ion-pair RP-HPLC. *Analytical biochemistry* 348, 243-251
18. Bradbury, A., *et al*. (2003) Antibodies in proteomics II: screening, high-throughput characterization and downstream applications. *Trends in biotechnology* 21, 312-317
19. Lee, H., *et al*. (2008) Chip-NMR biosensor for detection and molecular analysis of cells. *Nature medicine* 14, 869-874
20. Rudnitskaya, A., and Legin, A. (2008) Sensor systems, electronic tongues and electronic noses, for the monitoring of biotechnological processes. *Journal of industrial microbiology & biotechnology* 35, 443-451
21. Cannizzaro, C., *et al*. (2003) On-line biomass monitoring of CHO perfusion culture with scanning dielectric spectroscopy. *Biotechnology and bioengineering* 84, 597-610
22. Soons Z.I.T.A., *et al*. (2008) Online automatic tuning and control for fed-batch cultivation, *Bioprocess and Biosystems Engineering* 31(5), 453-467
23. Soons Z.I.T.A., *et al*. (2008) Observer design and tuning for biomass growth and kLa using online and offline measurements, *J of Process Control* 18, 621-631

24. Zhang, Y., and Edgar, T. (2006) Bio-reactor monitoring with multiway PCA and Model Based PCA. In *AIChE 2006 Anual Meeting*
25. Kasprow, R.P., *et al*. (1998) Correlation of fermentation yield with yeast extract composition as characterized by near-infrared spectroscopy. *Biotechnology progress* 14, 318-325
26. Thalen, M., *et al*. (2008) Improving the cellular pertussis vaccine: Increased potency and consistency. *Vaccine* 26, 653-663
27. Cunha, C.C.F., *et al*. (2002) An assessment of seed quality and its influence on productivity estimation in an industrial antibiotic fermentation. *Biotechnology and bioengineering* 78, 658-669
28. Zwietering, M.H., *et al*. (1992) Comparison of definitions of the lag phase and the exponential phase in bacterial growth. *The Journal of applied bacteriology* 72, 139-145
29. Robinson, T.P., *et al*. (2001) The effect of inoculum size on the lag phase of Listeria monocytogenes. *International journal of food microbiology* 70, 163-173
30. Robinson, T.P., *et al*. (1998) The effect of the growth environment on the lag phase of Listeria monocytogenes. *International journal of food microbiology* 44, 83-92
31. Yin, J., *et al*. (2007) Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *J Biotechnol* 127, 335-347
32. Walsh, G., and Jefferis, R. (2006) Post-translational modifications in the context of therapeutic proteins. *Nature biotechnology* 24, 1241-1252
33. Walsh, G. (2003) Biopharmaceutical benchmarks--2003. *Nature biotechnology* 21, 865-870
34. Graumann, K., and Premstaller, A. (2006) Manufacturing of recombinant therapeutic proteins in microbial systems. *Biotechnology journal* 1, 164-186
35. Yee, L., and Blanch, H.W. (1992) Recombinant protein expression in high cell density fed-batch cultures of Escherichia coli. *Bio/technology (Nature Publishing Company)* 10, 1550-1556
36. Hewitt, C.J., *et al*. (1999) The use of multi-parameter flow cytometry to compare the physiological response of Escherichia coli W3110 to glucose limitation during batch, fed-batch and continuous culture cultivations. *J Biotechnol* 75, 251-264
37. Hewitt, C.J., *et al*. (1999) Use of multi-staining flow cytometry to characterise the physiological state of Escherichia coli W3110 in high cell density fed-batch cultures. *Biotechnology and bioengineering* 63, 705-711
38. Gadgil, M., *et al*. (2005) Transcriptional response of Escherichia coli to temperature shift. *Biotechnology progress* 21, 689-699
39. Maurer, L.M., *et al*. (2005) pH regulates genes for flagellar motility, catabolism, and oxidative stress in Escherichia coli K-12. *Journal of bacteriology* 187, 304-319
40. Steel, R., and Maxon, W.D. (1966) Dissolved oxygen measurements in pilot- and production-scale novobiocin fermentations. *Biotech and Bioeng* 8, 97-108
41. Baneyx, F. (1999) Recombinant protein expression in Escherichia coli. *Current opinion in biotechnology* 10, 411-421
42. Lee, S.Y. (1996) High cell-density culture of Escherichia coli. *Trends Biotechnol* 14, 98-105
43. Sorensen, H.P., and Mortensen, K.K. (2005) Advanced genetic strategies for recombinant protein expression in Escherichia coli. *J Biotechnol* 115, 113-128
44. Pugsley, A.P. (1993) The complete general secretory pathway in gram-negative bacteria. *Microbiological reviews* 57, 50-108
45. Fahnert, B., *et al*. (2004) Inclusion bodies: formation and utilisation. *Advances in biochemical engineering/ biotechnology* 89, 93-142

46. Chou, C.P. (2007) Engineering cell physiology to enhance recombinant protein production in Escherichia coli. *Applied microbiology and biotechnology* 76, 521-532
47. Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiological reviews* 60, 512-538
48. Mason, C.A., *et al*. (1999) Heat-induced expression and chemically induced expression of the Escherichia coli stress protein HtpG are affected by the growth environment. *Applied and environmental microbiology* 65, 3433-3440
49. Kaiser, C., *et al*. (2008) Applications of PAT - Process Analytical Technology in Recombinant Protein Processes with Escherichia coli. *Eng. Life Sci.* 8, 7
50. Gnoth, S., *et al*. (2007) Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control. *J Biotechnol*
51. Hensing, M.C., *et al*. (1995) Physiological and technological aspects of large-scale heterologous-protein production with yeasts. *Antonie van Leeuwenhoek* 67, 261-279
52. van Dijken, J.P., *et al*. (1993) Kinetics of growth and sugar consumption in yeasts. *Antonie van Leeuwenhoek* 63, 343-352
53. Fukuhara, H. (2003) The Kluyver effect revisited. *FEMS yeast research* 3, 327-331
54. Werner, R.G., *et al*. (2007) Glycosylation of therapeutic proteins in different production systems. *Acta Paediatr Suppl* 96, 17-22
55. Wildt, S., and Gerngross, T.U. (2005) The humanization of N-glycosylation pathways in yeast. *Nature reviews* 3, 119-128
56. van Wezel, A.L. (1983) Microcarrier technology. Present status and prospects. *Developments in biological standardization* 55, 3-9
57. Molowa, D.T., and Mazanet, R. (2003) The state of biopharmaceutical manufacturing. *Biotechnology annual review* 9, 285-302
58. Browne, S.M., and Al-Rubeai, M. (2007) Selection methods for high-producing mammalian cell lines. *Trends Biotechnol* 25, 425-432
59. Griffin, T.J., *et al*. (2007) Advancing mammalian cell culture engineering using genome-scale technologies. *Trends Biotechnol* 25, 401-408
60. Barnes, L.M., *et al*. (2006) Phenotypic variation during cloning procedures: analysis of the growth behavior of clonal cell lines. *Biotechnology and bioengineering* 94, 530-537
61. Jenkins, N., *et al*. (2008) Post-translational modifications of recombinant proteins: significance for biopharmaceuticals. *Molecular biotechnology* 39, 113-118
62. Butler, M. (2006) Optimisation of the cellular metabolism of glycosylation for recombinant proteins produced by Mammalian cell systems. *Cytotechnology* 50, 57-76
63. Baldi, L., *et al*. (2007) Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. *Biotechnology letters* 29, 677-684
64. Kost, T.A., *et al*. (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nature biotechnology* 23, 567-575

Summary

Process Analytical Technology or PAT is an initiative launched by the American Food and Drug Administration (FDA) in 2003 and has been adopted by the European Medicines Agency (EMEA) and the International Conference for Harmonization (ICH) as well. By this initiative, the regulators want to overcome manufacturing rigidity caused by the regulations that govern the pharmaceutical industry and move to more flexible processing. This can be achieved by providing scientific evidence of the ranges of processing conditions and -parameters (critical process attributes) that will result in the manufacture of products with the desired specifications (critical product attributes). The interaction between critical process and product attributes needs to be understood. These scientifically supported manufacturing ranges constitute the so called process design space.

In addition to this, on line measurements (using PAT tools and sensors) need to monitor the process in order to assure that the process runs within the specified ranges of critical process attributes, thus within the process design space. By applying this principle, the assurance of product quality can rely less on the mandatory end product testing (as is now usually the case) and more on the actual process data. When all what is critical is measured during the process and the tolerance of these critical attributes is known and understood, quality assurance can actually become part of the process itself.

The increased scientific understanding allows more flexibility in manufacturing, because deviations from the target process can be assessed for their impact on product quality, rather than rejection of the batch or only accepting it after testing of the end product. Also process optimizations and subsequent changes of the process target settings within the process design space are no longer subject to regulatory approval when the process design space has previously been approved for registration.

FDA's original PAT guidance document was intended for small pharmaceutical molecules. Larger molecules (i.e. therapeutic proteins or antibodies) or vaccines (i.e. inactivated viruses or bacteria, or parts of them) are more complex and therefore application of PAT is considered more difficult. Especially the cultivation process step that is usually part of the processes for the manufacture of these biological products is a complex unit operation. It involves the growth of cells or micro organisms that express the protein of interest using a bioreactor system. In the case of whole cell vaccines, it is the cultivated pathogen that is the actual product. These processes and their underlying mechanisms are much more complex than the processes for small pharmaceutical molecule, where chemical synthesis is the product forming unit operation.

In this thesis the application of PAT on a complex biological product, a whole cell vaccine against whooping cough disease, is investigated. A whole cell vaccine means that the product consists of inactivated whole cells of the pathogen, in this case the bacterium *Bordetella pertussis*. The vaccine functions by inducing a protective immune response against proteins on the outer membrane that are presented to the patient's immune system. The composition of the outer membrane of the bacterium can therefore be considered crucial for vaccine efficacy and product quality. Many of the outer membrane proteins are involved in bacterial virulence. They play a role in the attachement to human epithelial cells and subsequent colonization and infection.

In order to be able to asses the outer membrane composition, a DNA microarray analysis was developed to measure mRNA levels of all the genes of the bacterial cells during cultivation. By measuring mRNA expression levels on samples taken during processing, understanding was built on how process conditions influence growth and outer membrane composition and what this means for the expected quality of the final vaccine. Fifty-six conserved marker genes were identified to be activated when *Bordetela pertussis* enters its virulent state. These include all genes that code for outer membrane proteins involved in inducing a protective immune response. The expression levels of these marker genes could then be used to calculate a quantitative product quality score. This score is a weighed indication of *Bordetella pertussis*' virulent state and can thus be expected to indicate product quality during cultivation.

In this way the response of the bacteria to changes during normal and disturbed processing could be quantified. For instance, the response of the cells and the expected effect on product quality after a period of oxygen depletion or the depletion of nutrients at the end of the process were determined. Oxygen limitation was shown to be reversible in overall gene expression levels and did not affect the virulence marker genes at the end of cultivation. Also by determining the product quality score at many time points during cultivation, the optimal point of harvest was determined. The optimal harvest point proved to be dependent on the concentrations of the nutrients lactate and glutamate.

Because the DNA microarray analysis is an off line measurement, it can not be used as PAT tool. In fact, it is difficult to measure the entire composition of the outer membrane by an on line PAT sensor. To overcome this, on line near infrared (NIR) spectroscopy was implemented. NIR can be used in two ways. First to measure fingerprints of the entire process by determining the variance between many spectra taken during processing, which allows batch-to-batch comparison. Second to measure specific process parameters on line that were previously only available off line, such as optical density or nutrient concentrations.

Finally, in order to determine the process design space, an experimental design (DoE) was followed. The ranges of process conditions that allow the cultivation of *Bordetella pertussis* with the desired outer membrane composition were explored, which was confirmed at the end of each batch using DNA microarrays. During these cultivations all process data, including NIR, were collected in a central database. These data were subsequently processed to build a statistical model that encompasses all the explored ranges and thus describing the process design space. This model can be used to check if new production runs fall inside the explored process design space, so that they can be expected to lead to the desired product quality. This approach demonstrates how quality assurance can be built into a bacterial cultivation process for vaccine manufacturing, thus fulfilling the PAT principles.

Samenvatting

Process Analytical Technology oftewel PAT is een initiatief dat in 2003 door de Amerikaanse Food and Drug Administration (FDA) is gestart. Inmiddels is het ook door Europese regelgevers overgenomen en is het opgenomen in richtlijnen van de International Conference of Harmonization (ICH). Door dit initiatief probeert de regelgever de starheid van (bio)farmaceutische processen te doorbreken en meer flexibiliteit toe te staan. De starheid wordt veroorzaakt door de strenge regelgeving die voor deze industrie van toepassing is. Meer flexibiliteit kan alleen worden toegestaan als er voldoende wetenschappelijk bewijs is dat kritieke procescondities binnen een bepaald bereik resulteren in een product van voldoende kwaliteit. De invloed van procesparameters op (de vorming van) het product tijdens het proces dient hierbij volledig bekend te zijn. Het gecombineerde onderbouwde bereik van alle procesparameters samen vormt als het ware een multi-dimensionale ruimte, waarbinnen het proces altijd een product van de gewenste kwaliteit oplevert. Deze ruimte wordt de *process design space* genoemd.

Daarnaast dienen on line metingen tijdens het proces (met zogenaamde PAT tools en –sensoren) het proces goed te volgen zodat kan worden gecontroleerd of het proces nog binnen het gespecificeerde bereik van de respectievelijke kritische procesparameters verloopt. Deze toepassing leidt ertoe dat al tijdens het proces uitspraken over de kwaliteit van het product kunnen worden gedaan. Immers, wanneer de PAT tools hebben aangetoond dat het proces het gewenste verloop heeft gehad en in ieder geval binnen de *design space* is gebleven, is de productkwaliteit geborgd. Hierdoor hoeft men minder op het testen van het product na afloop van het proces te vertrouwen, hetgeen nu de verplichting is, maar kan men meer van de on line procesdata uitgaan. Indien alle kritische parameters daadwerkelijk tijdens het proces kunnen worden gemeten en de tolerantie op deze parameters in termen van productkwaliteit bekend is, dan wordt kwaliteitsborging als het ware onderdeel van het proces zelf.

Het verhoogde wetenschappelijk inzicht in de interacties tussen product en proces staan meer flexibiliteit in de uitvoering van het proces toe, doordat afwijkingen in het proces kunnen worden ingeschat op hun effect op de kwaliteit van het product. Met andere woorden, of ze binnen of buiten de *design space* vallen. Wanneer het proces binnen de *design space* blijft, hoeft het product niet te worden afgekeurd of de test op het eindproduct te worden afgewacht. Daarnaast zijn veranderingen van het proces binnen de *design space* als uitkomst van optimalisatie studies ook toegestaan. Deze hoeven niet ter goedkeuring aan de regelgever te worden voorgelegd, wanneer de hele *design space* al eerder door de regelgever is goedgekeurd.

De originele PAT richtlijn van de FDA was eigenlijk alleen bedoeld voor kleine, chemische farmaceutische moleculen. Grotere moleculen (zoals therapeutische eiwitten of

antilichamen) of vaccins (vaak geinactiveerde bacteriën of virussen, of onderdelen daarvan) zijn veel complexer. Daarom kan toepassing van PAT op de productieprocessen voor deze biofarmaceutica als ingewikkelder worden beschouwd. Met name de cultivatie stap, waarbij dierlijke cellen, gisten of bacteriën worden opgekweekt, is een zeer complexe stap. Het is meestal de eerste stap in een biofarmaceutisch productieproces en het is de stap waarbij het eigenlijke product wordt gevormd. Of, in het geval van vaccins, waarbij het gekweekte pathogeen zelf het product is. Deze processtap is veel complexer dan de chemische synthese die de product-vormende stap voor kleine chemische geneesmiddelen vormt.

In dit proefschrift worden de mogelijkheden voor toepassing van PAT op een complex biofarmaceuticum, een cellulair vaccin tegen kinkhoest, onderzocht. Een cellulair bacterieel vaccin betekent dat het product uit geïnactiveerde cellen van het pathogeen bestaat, in dit geval de bacterie *Bordetella pertussis*. De werking van een dergelijk klassiek vaccin is gebaseerd op het opwekken van een immuunrespons tegen de afgedode bacteriën die bescherming biedt tegen infectie met levende bacteriën. Met name de buitenkant van de bacteriële cel wordt herkend door het immuun systeem en de immunologische afweer is dan ook gericht tegen structuren op de buitenkant van de cel. De samenstelling van de buitenmembraan is dus van groot belang voor de effectiviteit van het vaccin. Veel van de eiwitten op de buitenmembraan zijn betrokken bij de virulentie van de bacterie. Ze spelen bijvoorbeeld een rol bij de hechting van de cel aan een gastheercel en het verder koloniseren en infecteren van de keelholte.

Om uitspraken over de samenstelling van de buitenmembraan te kunnen doen werden de expressie niveaus van alle genen van de bacteriële cel tijdens het cultivatieproces gemeten met DNA microarrays. Uit deze metingen volgde informatie over het effect van procescondities op de expressie van de verschillende genen, waaronder de genen die voor de buitenmembraan-eiwitten coderen. Op deze manier kon het effect van veranderende procescondities op de verwachte kwaliteit van het product worden ingeschat. Uiteindelijk werden 56 marker genen gevonden die allen geactiveerd worden als *Bordetella pertussis* virulent wordt. Hierbij zaten alle genen die coderen voor (buitenmembraan)eiwitten waarvan bekend is dat ze betrokken zijn bij het induceren van een beschermende immuunrespons na vaccinatie. Het expressie niveau van deze genen werd omgerekend naar een gewogen kwantitatieve score voor productkwaliteit waarmee de verwachte productkwaliteit tijdens het kweekproces kon worden aangegeven.

Met deze score kon het effect van veranderingen tijdens het proces en als gevolg van procesverstoringen worden gemeten. Bijvoorbeeld het effect van een verstoring van de zuurstofvoorziening en het effect van lage concentraties van nutriënten aan het einde van de kweek zijn onderzocht. Zuurstof limitatie bleek een reversibel effect te hebben op de expressie van alle genen en beïnvloedde de expressie van de virulentie genen aan het einde van de kweek niet. Door tijdens een groot aantal tijdspunten tijdens het kweekproces de productkwaliteitsscore te bepalen kon het optimale oogstmoment worden bepaald. Dit optimale oogstmoment bleek sterk afhankelijk van de concentraties van de nutrienten lactaat en glutamaat aan het einde van het kweekproces.

De DNA microarrays zijn een bewerkelijke off line analyse en kunnen daarom niet gebruikt worden als PAT tool voor online metingen aan het proces. Echter, er zijn geen sensoren beschikbaar die tijdens het proces metingen kunnen doen aan de samenstelling van de buitenmembraan. Om dit probleem op te lossen is on line nabij infrarood (NIR) spectroscopie toegepast. Deze sensor kan op twee manieren worden gebruikt. Enerzijds om een zogenaamde vingerafdruk van het gehele proces te maken door de variatie tussen spectra, genomen gedurende de gehele batch, te bepalen. Hiermee kunnen verschillende batches met elkaar worden vergeleken. Anderzijds door specifieke kritische procesparameters te meten, die voorheen alleen off line beschikbaar waren, zoals nutriënt concentraties of de optische dichtheid van de cultuur.

Uiteindelijk werd een experiment ontworpen volgens de DoE methode waarbij de *process design space* voor de kweek van *Bordetella pertussis* in kaart werd gebracht. De verschillende kritische procesparameters werden over een bepaald bereik getest waarbinnen goede productkwaliteit kon worden verwacht. In een DoE ontwerp worden alle parameters tegelijkertijd gevarieerd waardoor de interactie tussen de parameters ook zichtbaar wordt. De productkwaliteit werd aan het eind van elke cultivatie bepaald met de microarray analyse. Alle procesgegevens van deze kweken, inclusief de NIR metingen, werden in een centrale databank opgeslagen. Vervolgens werden deze gegevens gebruikt om een statistisch model te construeren dat het verkende bereik van alle kritische procesparameters omvat. Dit model beschrijft de *process design space*. Het model kan gebruikt worden tijdens nieuwe kweekprocessen om te controleren of het proces binnen de *design space* valt en of de bacteriën dus de gewenste samenstelling van de buitenmembraan hebben. Hiermee is voor een complex biologisch proces voor vaccinproductie aangetoond hoe kwaliteitsbewaking onderdeel van het proces zelf kan worden en dat dus kan worden voldaan aan de principes van PAT.

Dankwoord

Het laatste en veelal meestgelezen hoofdstuk van een proefschrift is natuurlijk opgedragen aan alle mensen die dank zijn verschuldigd voor hun hulp, steun, kritiek, humor, relativering of onvoorwaardelijke liefde. Het zijn deze mensen die dit boekje mogelijk hebben gemaakt.

Allereerst mijn (co)promotoren. Hans, dank voor je immer terechte commentaar, je complimenten en bemoedigende woorden. Het is een eer en genoegen om jouw laatste promovendus te zijn in een lange reeks die bij RIVM/NVI en Proceskunde zijn gepromoveerd. Dirk, dank voor je onaflatende en altijd constructieve kritiek. Waar ik dacht dat het wel goed genoeg was, vond jij altijd punten waar verbetering mogelijk was. Met name tijdens het schrijven van het laatste hoofdstuk wisselden we dagelijks nieuwe versies uit, waarvan elke steeds beter dan de vorige was. Leo, dank voor de ruimte en vrijheid die je me in de laatste fase van mijn promotie hebt gegeven. Nooit heb je bezwaar gemaakt tegen de uren die in dit proefschrift zitten, ook nadat het PaRel project was afgelopen. Je bent een “PAT believer” geworden en verschillende elementen uit dit proefschrift zijn al geïntegreerd in de manier van werken binnen PO. Tenslotte Coen, geestelijk vader van het PaRel project en een echte mentor tijdens de promotie. Na de hectiek van het eerste jaar, heeft het PaRel project bij alle partners tot geweldige resultaten geleid. Jouw visie, enthousiastme en onvoorwaardelijke steun zijn een heel belangrijke motivatie voor mij geweest. Het is geweldig om iemand te hebben die in je gelooft. We zullen elkaar in de toekomst zeker nog vaak ’s avonds bellen.

Dan mijn twee paranimfen Bas en Paul: de ruggegraat van het PaRel team. Ik neem mijn hoed af en buig diep voor jullie inzet tijdens het project. Paul, jij was er vanaf het begin bij. De discussies met Siemens over de samenhang met het BIOCS project en het incasseren van vertraging op vertraging staan mij nog helder voor de geest. Je bent tegen wil en dank expert in PCS7 geworden en altijd bereid om het gecrashte systeem weer op te lappen. ’s Nachts aanenten was ook nooit een probleem, zelfs toen je daarvoor op de fiets uit Vianen moest komen met een lekke band op de koop toe. We zijn samen opgegroeid met het PAT gedachtengoed en ik ken niemand die de technische consequenties van het ontwerpen van een PAT applicatie beter overziet dan jij. Bas, ik zie je nog zitten tijdens je sollicitatie gesprek, waarbij met Gino een discussie loskwam over je onderkomen in Carol’s Paradise. Je bent een zeldzame combinatie van gezelligheid en gedrevenheid en zonder jouw bijdrage met de microarrays had hier slechts een heel mager proefschrift gelegen. Het is voor mij nog steeds verbazingwekkend om te zien hoe jij van een onwaarschijnlijke chaoot in een ongelooflijke perfectionist verandert als het op experimenteren aankomt. Ik kan je alleen zeggen dat zo’n promotie een hell of a job is, maar ik weet zeker dat jij het ook kan. Ga ervoor de komende jaren en ik zal je helpen waar ik kan. Noblesse oblige.

Dan waren er nog mijn studenten die veel werk voor me hebben verzet. Paloma, jij was de eerste en dat was voor mij wel een vuurdoop vanwege je geweldige temperament. Hester, jij bewees dat vrouwen gewoon beter in het lab zijn dan mannen. Bedankt voor het leggen van de basis voor mijn proefschrift in hoofdstuk 2. Joeri, jij was met afstand de gezelligste student, maar dat ging zeker niet ten koste van je inzet. We zullen zeker nog afspreken voor een biertje. Ruud "de Roller", met afstand de meest onzichtbare student, bedankt voor je werk aan de kweek pH. Jun, thanks for giving me insight in the Chinese culture and for your wonderfully warm personality. Don't become too attached to Holland, else you cannot "Go back China" anymore. Good luck with your thesis.

Verder zijn er mensen die hun heil elders zoeken. Gino, dank voor je enhousiasme en gevoel voor humor. Je haast Spartaanse huishouden in Bilthoven heeft plaatsgemaakt voor warm familieleven. Ik weet binnenkort hoe goed dit bevalt. Heel veel succes in het Gentse. Zita, ik hoop dat je herstel blijvend is en wens je alle goeds toe in Portugal. Onze samenwerking was soms stroef, maar van wrijving komt glans. Jij, en ik wil hier Ton ook zeker niet vergeten, hebt een geweldige bijdrage aan het project geleverd. Jeroen, je was ook altijd een gezellige collega op het lab. Marcel, eerst als collega, later als afdelingshoofd, maar met name als de ontwikkelaar van het THIJS medium, waar al het kweekwerk in dit proefschrift mee is uitgevoerd. En natuurlijk Jan… Je kon gelukkig niet lang ontkennen dat het NVI een heerlijke plek is om je kweekjes te draaien. En voor bakkie doen!

Ook alle andere mensen van de unit O&O van het NVI en dan speciaal de afdeling (L)PO, wil ik hierbij bedanken. Met name Gideon voor zijn steun voor het PaRel project in de beginfase, Bert voor het analyseren van al mijn saaie samples en Claire voor je aanstekelijke enthousiastme; fijn dat je in mijn commissie wilde plaatsnemen. Karin, dank voor je hulp en advies bij de "papieren" kant van het PaRel project. En natuurlijk Wilma voor de geweldige layout van dit boekje. Jij hebt me echt uren frustratie en computerleed bespaard en het is dankzij jou een prachtig proefschrift geworden. Alle andere collega's ontzettend bedankt voor de goede sfeer en de vele gezellige momenten in het lab of de koffiekamer.

Buiten O&O wil ik met name Marvin en Gerhan bedanken voor de samenwerking in het begin van het project. De vergaderingen bij Applikon en Siemens staan mij nog levendig voor de geest. Deze bedrijven bleken geweldige projectpartners en de samenwerking als geheel is met de nodige ups en downs al met al toch altijd zeer prettig geweest. Arthur, ik herinner me de avondsessie bij Applikon nog goed waarbij we de subsidieaanvraag hebben afgemaakt en uiteindelijk al de hele basis voor het PaRel project hebben gelegd. Jaap, bedankt voor de samenwerking, met name tijdens de vele stuurgroep vergaderingen waarbij we steevast over de uitputting van de subsidie moesten steggelen. Ingrid, je bent tijdens dit project uitgegroeid

tot een ware PAT discipel en je bent in staat de evidente voordelen op een geweldige manier uit te dragen. Ik zal de manier waarop je het abstracte PAT principe kan overbrengen nooit vergeten. Veel succes in je nieuwe job. Leo, een verkoper zal ik nooit worden, maar ik heb wel geleerd van de manier waarop jij tegen zaken aankijkt. Elsa en Alexander, bedankt voor het uiteindelijk afmaken van de BIOCS software library waardoor wij verder konden met onze experimenten. Hans voor de goede samenwerking en je geweldig nuchtere en zakelijke manier van redeneren, ik vond het ook erg leuk je in Beijing te treffen; veel plezier en succes in China. Erik als geweldige opvolger van Hans. Nog bedankt dat je me op sleeptouw hebt genomen tijdens de IFPAC. Jan voor de ondersteuning van bij het debuggen van PCS7 en Kjell idem voor SIPAT. Geniet van je prachtige dochter. En verder alle andere mensen bij Siemens en Applikon die ik ben vergeten, maar die er wel voor gezorgd hebben dat de experimenten uit de laatse hoofdstukken van dit proefschrift konden worden uitgevoerd.

En dan was er nog mijn tweede werkplek, de vakgroep proceskunde in Wageningen. Ik was er bijna nooit, maar ik vond het altijd leuk om er te zijn. Met name als er iets te vieren viel of als er een reisje in het verschiet lag, was ik er graag bij. De eerste echte kennismaking was de brainstormsessie eind 2005, waarbij Hedy actief beleid voerde om het alcoholgebruik op te voeren. Franske werkte ook goed mee. Joost, ik heb jou daar als partner in crime ontmoet en onze sterke voorliefde voor nog een laatste biertje heeft onze reputatie bestendigd in Kopenhagen. Lades was duidelijk the place to be, alhoewel niet iedereen dat met ons eens was. Packo, Marieke, Koen, Eduard, Dorinde, Annette, Floor en vele anderen heb ik leren kennen op die reis. In Japan werd de afwezigheid van Joost ruimschoots gecompenseerd, door Klaske, Sina, Maria en eenmalig René, thanks for the wonderful evenings in Japanese restaurants bars and disco's. Oh, en de presentaties waren natuurlijk ook zeer leerzaam. Verder wil ik ook alle andere mensen van de vakgroep bedanken voor de geweldige sfeer die ik mocht proeven tijdens de gelegenheden dat ik met jullie mee mocht.

Aan het begin van mijn promotie had ik nog wel eens wat vrije tijd en die breng ik graag door met mijn vrienden. Mijn oude studiegenoten waarmee naar hartelust verhalen van "vroeger" kunnen worden opgehaald gaven altijd weer nieuwe energie. Thijs, Werner, Valesca, Erik, Marieke, Joke, ik heb jullie leren kennen in de mooiste tijd van mijn leven, en dat krijg ik bij jullie altijd weer een beetje terug. Ook alle anderen die ik bij gelegenheid nog wel eens tref met een biertje in de hand, bedankt voor alle gezelligheid.
"Vroeger" gaat eigenlijk ook al op voor mijn lieve vrienden uit Ridderkerk. Raymond, Bas, Marjolein, Femke, Matthijs, Jolanda, Martijn, Sietske, Jan en An, onze vakanties in mei zijn elk jaar weer geweldig en geven het gevoel een maand weg te zijn geweest. Als er iets is dat we samen goed kunnen dan is het wel lachen. Het is altijd een feest om bij jullie te zijn.

Wim en Sjanie, dank voor jullie belangstelling en bemoedigende woorden. Jullie zijn geweldige schoonouders en sinds kort ook opa en oma. Geniet van deze bijzondere tijden.

Pap en mam, met dit boekje haal ik mijn laatste schooldiploma. Al zou ik verder willen, het kan niet meer; dit is het hoogst haalbare. Dankzij jullie steun kon ik de eerste doctor in de familie worden. Jullie hebben, soms terecht, soms niet, weleens getwijfeld aan mijn inzet, met name op het VWO en de studie in Leiden, maar nooit aan mij; jullie wisten dat ik het kon. Dank jullie wel voor alles; alle zorgen en alle keren dat jullie trots waren. Ik weet zeker dat jullie een fantastische opa en oma voor Bouke worden. We gaan een geweldige tijd tegemoet!

En dan als laatste, maar nooit als minste, mijn lieve Mariska. Teerbeminde. Een brok in de keel, zelfs tijdens het typen, als ik denk aan wat jij voor mij betekent. Je bent er altijd als het slecht met me gaat. Je danst met me mee als het goed met me gaat. Love is you. Love is me. Love is knowing we can be. Moeder van mijn zoon. Liefde van mijn leven. Ik hou van je.

Curriculum Vitae

Mathieu Streefland was born in Gouda (The Netherlands) on the 2nd of March 1978. After primary school in his home village Zwammerdam he received his secondary education at the Christelijk Lyceum, later Groene Hart Lyceum in Alphen aan den Rijn at which he graduated at the VWO level in 1996. That same year he started studying Biopharmaceutical Sciences at Leiden University. His two MSc theses were conducted at Leiden university. One focused on the role of p53 mediated apoptosis in atherosclerotic mice and the second on the role of the HPA axis and corticosterone in aggressive behaviour in rats. He received his MSc degree in 2001.

In 2002 he started working as interim professional at Yacht Technology. He was stationed at the sector Vaccines of the Dutch national institute of health and the environment (RIVM) to work on the development of a manufacturing process for a live attenuated vaccine against RSV and later the development of serum free cultivation media for animal cells. In 2003 the sector Vaccines joined with the public foundation SVM to form the Netherlands Vaccine Institute (NVI). Later that year, Mathieu joined in the writing of a grant proposal for the development of an advanced monitoring and control strategy for the cultivation process step for vaccine manufacturing: Parametric Release for Biopharmaceuticals, in short PaRel. After this grant was awarded in july 2003 he joined the PaRel project team. At the end of 2003 he was employed by the NVI as scientist.

In 2004 Mathieu became project leader of the PaRel project and after the promising initial results, the decision was made to write this PhD thesis on the subject of PAT application on a vaccine cultivation process step in collaboration with the bioprocess engineering group at Wageningen University.

Publications

Streefland M, van de Waterbeemd B, Happe H, van der Pol LA, Beuvery EC, Tramper J, Martens DE. 2007. PAT for vaccines: the first stage of PAT implementation for development of a well-defined whole-cell vaccine against whooping cough disease. Vaccine 25(16):2994-3000.

Van Sprang ENM, Streefland M, Ramaker HJ, Van der Pol LA, Beuvery EC, Smilde AK. 2007. Manufacturing Vaccines: An Illustration of Using PAT Tools for Controlling the Cultivation of *Bordetella pertussis*. Quality Engineering 19(4):373-384

Soons ZITA, Streefland M, Van Straten G, Van Boxtel AJB. Assessment of near infrared and "software sensor" for biomass monitoring and control. 2008. Chemometrics and Intelligent Laboratory Systems 94(2):166-174

Streefland M, van de Waterbeemd B, Kint J, van der Pol LA, Beuvery EC, Tramper J, Martens DE. 2009. Evaluation of a critical process parameter: oxygen limitation during cultivation has a fully reversible effect on gene expression of *Bordetella pertussis*. Biotechnology and Bioengineering 102(1):161-167.

Training activities

Conferences

PAT for Biologics (Washington DC, USA, 2004)
Advances in Microarray Technology (Londen, UK, 2005)
Bioproduction 2005 (Amsterdam, The Netherlands, 2005)
IFPAC 2006 (Washington DC, USA, 2006)
11e Nederlands Biotechnologie Congres (Ede, The Netherlands, 2006)
Bioproduction 2006 (Dublin, Ireland, 2006)
WCVII 2006 (Montréal, Canada, 2006)
European Conference on Biotechnology (Barcelona, Spain, 2007)
Heidelberg PAT Conference (Heidelberg, Germany, 2007)
Vaccines Asia 2007 (Beijing, China, 2007)
PAT and Quality By Design (Amsterdam, The Netherlands, 2008)
12e Nederlands Biotechnologie Congres (Ede, The Netherlands, 2008)
IRDG Voorjaarsdag (Geleen, The Netherlands, 2008)

Courses

VMT (NVI, 2007)
Management van Onderzoeksprojecten (HORIZON, 2005)
NVI GMP (NVI, 2005)
MBTI (NVI, 2003)
Inleiding DoE (The Production Factory, 2006)
Troubleshooting of Industrial Processes (TIPb, 2007)
Design of Experiments Pharma Applications (Umetrics, 2008)

Optionals

Process Engineering PhD study tour to Denmark and Sweden (2006)
Process Engineering PhD study tour to Japan (2008)
Preparation TS Grant Proposal Parametric Release for Biopharmaceuticals (2003)
Preparation EUREKA Grant Proposal PAT for Lyophylization (2008)

Layout: Wilma Witkamp

Kaft: Cy5 fluorescentie signaal van een DNA microaray met het volledige genoom van *Bordetella pertussis*

Het onderzoek beschreven in dit proefschrift is uitgevoerd op het Nederlands Vaccin Instituut te Bilthoven, Nederland